

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されて
る事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office.

出願年月日
Date of Application:

2003年 6月27日

出願番号
Application Number:

特願2003-183734

[ST. 10/C]:

[JP 2003-183734]

願人
Applicant(s):

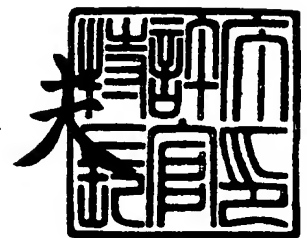
株式会社日立製作所

CERTIFIED COPY OF
PRIORITY DOCUMENT

2004年 5月24日

特許庁長官
Commissioner,
Japan Patent Office

今井康夫



BEST AVAILABLE COPY

【書類名】 特許願

【整理番号】 K03009331A

【あて先】 特許庁長官殿

【国際特許分類】 G06F 3/06

【発明者】

【住所又は居所】 神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 R A I D システム事業部内

【氏名】 平川 裕介

【発明者】

【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所 システム開発研究所内

【氏名】 荒川 敬史

【発明者】

【住所又は居所】 神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 R A I D システム事業部内

【氏名】 武田 貴彦

【発明者】

【住所又は居所】 神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 R A I D システム事業部内

【氏名】 佐藤 孝夫

【特許出願人】

【識別番号】 000005108

【氏名又は名称】 株式会社日立製作所

【代理人】

【識別番号】 100075096

【弁理士】

【氏名又は名称】 作田 康夫

【手数料の表示】

【予納台帳番号】 013088

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 記憶システム

【特許請求の範囲】

【請求項 1】

上位装置に接続され、前記上位装置との間でデータの送受を行う第一の記憶システムと、

前記第一の記憶システムに接続され、前記第一の記憶システムからデータを受け取る第二の記憶システムとを有し、

前記第一の記憶システムは、

前記上位装置から送られたデータが書き込まれる第一の記憶領域と、

前記第一の記憶領域に書き込まれたデータと前記データに関する更新情報とが書き込まれる第二の記憶領域とを有しており、

前記第二の記憶システムは、前記第二の記憶領域から読み出されたデータと前記データに関する更新情報とが格納される第三の記憶領域とを有するものであり

、
前記第三の記憶領域に格納されるデータ及び更新情報は、所定の時間間隔毎に、前記第一の記憶システムから読み出されるものであることを特徴とする記憶システム群。

【請求項 2】

請求項 1 に記載の記憶システム群において、

前記第二の記憶システムは、前記第二の記憶領域から読み出されたデータの個数に応じて、前記所定の時間間隔を決定するように制御するものであることを特徴とする記憶システム群。

【請求項 3】

請求項 1 に記載の記憶システム群において、

前記第二の記憶システムは、前記第一の記憶システムと前記第二の記憶システムとの間で送受されるデータの通信量に応じて、前記所定の時間間隔を決定するように制御するものであることを特徴とする記憶システム群。

【請求項 4】

請求項 1 記載の記憶システム群において、

前記第二の記憶システムは、前記第三の記憶領域が保持している記憶容量に応じて、前記所定の時間間隔を決定するように制御するものであることを特徴とする記憶システム群。

【請求項 5】

請求項 1 に記載の記憶システム群において、

前記第二の記憶システムは、前記第二の記憶システム自身の処理負荷に応じて、前記所定の時間間隔を決定するように制御するものであることを特徴とする記憶システム群。

【請求項 6】

請求項 1 に記載の記憶システム群において、

前記第二の記憶システムは、前記第一の記憶システム内の前記第二の記憶領域が保持している記憶容量についての情報を、前記第一の記憶システムから読み出し、前記読み出された情報に応じて、前記所定の時間間隔を決定するように制御するものであることを特徴とする記憶システム群。

【請求項 7】

請求項 1 に記載の記憶システム群において、

前記第二の記憶領域に書き込まれる更新情報は、前記第一の記憶領域に書き込まれるデータの更新順序に関するものであることを特徴とする記憶システム群。

【請求項 8】

請求項 7 に記載の記憶システム群において、

前記第一の記憶システムは、前記第二の記憶領域についての管理情報を所有するものであり、

前記第二の記憶システムは、前記第一の記憶システムが所有している前記第二の記憶領域についての管理情報を、前記第一の記憶システムから読み出し、前記読み出された管理情報に応じて、前記所定の時間間隔を決定するように制御するものであることを特徴とする記憶システム群。

【請求項 9】

請求項 1 に記載の記憶システム群において、

前記第一の記憶領域は、複数存在し、

前記第二の記憶領域に書き込まれる更新情報は、前記複数の第一の記憶領域に書き込まれるデータを対象にして作成されるものであることを特徴とする記憶システム群。

【請求項 10】

請求項 9 に記載の記憶システム群において、

前記第二の記憶領域に書き込まれる更新情報は、前記複数の第一の記憶領域に書き込まれるデータの更新順序に関するものであることを特徴とする記憶システム群。

【請求項 11】

請求項 9 に記載の記憶システム群において、

前記第一の記憶システムは、前記複数の第一の記憶領域に書き込まれるデータの更新情報の格納領域として前記第二の記憶領域を割当て、前記第二の記憶領域についての管理情報を所有するものであり、

前記第二の記憶システムは、前記第一の記憶システムが所有している前記第二の記憶領域についての管理情報を、前記第一の記憶システムから読み出し、前記読み出された管理情報に応じて、前記所定の時間間隔を決定するように制御するものであることを特徴とする記憶システム群。

【請求項 12】

請求項 1 に記載の記憶システム群において、

前記第二の記憶システムは、前記第二の記憶領域に格納されたデータ及び前記データに関する更新情報の送付を要求するコマンドを、前記所定の時間間隔毎に、前記第一の記憶システムに対して送信するものであり、

前記第一の記憶システムは、前記コマンドに応じて、前記第二の記憶領域に格納されたデータ及び前記データに関する更新情報を前記第二の記憶システムに対して送信するものであることを特徴とする記憶システム群。

【請求項 13】

請求項 1 に記載の記憶システム群において、

前記第二の記憶システムは、前記第一の記憶システム内の前記第一の記憶領域

に対応する第四の記憶領域を有しており、前記第三の記憶領域に格納されているデータ及び更新情報を基にして、前記第四の記憶領域に対してデータを格納することを特徴とする記憶システム群。

【請求項 14】

請求項 13 に記載の記憶システム群において、

前記第二の記憶システムは、前記第二の記憶システム自身の処理負荷に応じて、前記第四の記憶領域に対するデータの格納処理のタイミングを制御することを特徴とする記憶システム群。

【請求項 15】

請求項 1 に記載の記憶システム群において、

前記第一の記憶システムは、上位装置との間でデータを送受する上位アダプタと、前記上位アダプタで受けるデータを保存するキャッシュと、前記キャッシュに格納されたデータを転送するディスクアダプタと、前記ディスクアダプタの制御に応じてデータが格納される複数のディスクドライブとを有しており、

前記第一の記憶領域及び前記第二の記憶領域は、前記複数のディスクドライブ内の記憶領域から割当てられるものであることを特徴とする記憶システム群。

【請求項 16】

請求項 1 に記載の記憶システム群において、

前記第二の記憶システムは、上位装置との間でデータを送受する上位アダプタと、前記上位アダプタで受けるデータを保存するキャッシュと、前記キャッシュに格納されたデータを転送するディスクアダプタと、前記ディスクアダプタの制御に応じてデータが格納される複数のディスクドライブとを有しており、

前記第三の記憶領域は、前記複数のディスクドライブ内の記憶領域から割当てられるものであることを特徴とする記憶システム群。

【請求項 17】

上位装置に接続され、前記上位装置との間でデータの送受を行う第一の記憶システムと、

前記第一の記憶システムに接続され、前記第一の記憶システムからデータを受け取る第二の記憶システムと、

前記第二の記憶システムに接続され、前記第二の記憶システムからデータを受け取る第三の記憶システムとを有し、

前記第一の記憶システムは、前記上位装置から送られたデータが書き込まれる第一の記憶領域を有しており、

前記第二の記憶システムは、前記第一の記憶システム内の前記第一の記憶領域に書き込まれたデータと前記データに関する更新情報とが書き込まれる第二の記憶領域とを有しており、

前記第三の記憶システムは、前記第一の記憶システム内の前記第二の記憶領域から読み出されたデータと前記データに関する更新情報とが格納される第三の記憶領域とを有するものであり、

前記第三の記憶領域に格納されるデータ及び更新情報は、所定の時間間隔毎に、前記第二の記憶システムから読み出されるものであることを特徴とする記憶システム群。

【請求項 1 8】

請求項 1 7 に記載の記憶システム群において、

前記第三の記憶システムは、前記第二の記憶領域から読み出されたデータの個数に応じて、前記所定の時間間隔を決定するように制御するものであることを特徴とする記憶システム群。

【請求項 1 9】

請求項 1 7 に記載の記憶システム群において、

前記第三の記憶システムは、前記第二の記憶システムと前記第三の記憶システムとの間で送受されるデータの通信量に応じて、前記所定の時間間隔を決定するように制御するものであることを特徴とする記憶システム群。

【請求項 2 0】

請求項 1 7 に記載の記憶システム群において、

前記第三の記憶システムは、前記第三の記憶領域が保持している記憶容量に応じて、前記所定の時間間隔を決定するように制御するものであることを特徴とする記憶システム群。

【請求項 2 1】

請求項 1 7 に記載の記憶システム群において、

前記第三の記憶システムは、前記第三の記憶システム自身の処理負荷に応じて、前記所定の時間間隔を決定するように制御するものであることを特徴とする記憶システム群。

【請求項 2 2】

請求項 1 7 に記載の記憶システム群において、

前記第三の記憶システムは、前記第二の記憶システム内の前記第二の記憶領域が保持している記憶容量についての情報を、前記第二の記憶システムから読み出し、前記読み出された情報に応じて、前記所定の時間間隔を決定するように制御するものであることを特徴とする記憶システム群。

【請求項 2 3】

請求項 1 7 に記載の記憶システム群において、

前記第二の記憶領域に書き込まれる更新情報は、前記第二の記憶領域に書き込まれるデータの更新順序に関するものであることを特徴とする記憶システム群。

【請求項 2 4】

請求項 2 3 に記載の記憶システム群において、

前記第二の記憶システムは、前記第二の記憶領域についての管理情報を所有するものであり、

前記第三の記憶システムは、前記第二の記憶システムが所有している前記第二の記憶領域についての管理情報を、前記第二の記憶システムから読み出し、前記読み出された管理情報に応じて、前記所定の時間間隔を決定するように制御するものであることを特徴とする記憶システム群。

【請求項 2 5】

請求項 1 7 に記載の記憶システム群において、

前記第三の記憶システムは、前記第二の記憶領域に格納されたデータ及び前記データに関する更新情報の送付を要求するコマンドを、前記所定の時間間隔毎に、前記第二の記憶システムに対して送信するものであり、

前記第二の記憶システムは、前記コマンドに応じて、前記第二の記憶領域に格納されたデータ及び前記データに関する更新情報を前記第三の記憶システムに対

して送信するものであることを特徴とする記憶システム群。

【請求項 2 6】

請求項 1 7 に記載の記憶システム群において、

前記第三の記憶システムは、前記第一の記憶システム内の前記第一の記憶領域に対応する第四の記憶領域を有しており、前記第三の記憶領域に格納されているデータ及び更新情報を基にして、前記第四の記憶領域に対してデータを格納することを特徴とする記憶システム群。

【請求項 2 7】

請求項 2 6 に記載の記憶システム群において、

前記第三の記憶システムは、前記第三の記憶システム自身の処理負荷に応じて、前記第四の記憶領域に対するデータの格納処理のタイミングを制御することを特徴とする記憶システム群。

【請求項 2 8】

請求項 1 7 に記載の記憶システム群において、

前記第一の記憶システムは、前記上位装置から送られたデータを前記第一の記憶領域に書き込んだ場合、前記第一の記憶領域に書き込まれたデータを前記第二の記憶システムに対して送信するものであり、

前記第二の記憶システムは、前記第一の記憶システム内の前記第一の記憶領域に書き込まれたデータと前記データに関する更新情報を前記第二の記憶領域に書き込むものであることを特徴とする記憶システム群。

【請求項 2 9】

上位装置に接続され、前記上位装置との間でデータの送受を行う第一の記憶システムと、

前記第一の記憶システムに接続され、前記第一の記憶システムからデータを受け取る第二の記憶システムと、

前記第二の記憶システムに接続され、前記第二の記憶システムからデータを受け取る第三の記憶システムとを有し、

前記第一の記憶システムは、前記上位装置から送られたデータが書き込まれる第一の記憶領域を有しており、

前記第二の記憶システムは、前記第一の記憶システムから送られたデータが書き込まれる第二の記憶領域と、前記第二の記憶領域に書き込まれたデータと前記データに関する更新情報とが書き込まれる第三の記憶領域とを有しており、

前記第三の記憶システムは、前記第二の記憶システム内の前記第三の記憶領域から読み出されたデータと前記データに関する更新情報とが格納される第四の記憶領域とを有するものであり、

前記第四の記憶領域に格納されるデータ及び更新情報は、所定の時間間隔毎に、前記第三の記憶システムから読み出されるものであることを特徴とする記憶システム群。

【請求項 3 0】

請求項 2 9 に記載の記憶システム群において、

前記第三の記憶システムは、前記第三の記憶領域から読み出されたデータの個数に応じて、前記所定の時間間隔を決定するように制御するものであることを特徴とする記憶システム群。

【請求項 3 1】

請求項 2 9 記載の記憶システム群において、

前記第三の記憶システムは、前記第二の記憶システムと前記第三の記憶システムとの間で送受されるデータの通信量に応じて、前記所定の時間間隔を決定するように制御するものであることを特徴とする記憶システム群。

【請求項 3 2】

請求項 2 9 記載の記憶システム群において、

前記第三の記憶システムは、前記第四の記憶領域が保持している記憶容量に応じて、前記所定の時間間隔を決定するように制御するものであることを特徴とする記憶システム群。

【請求項 3 3】

請求項 2 9 に記載の記憶システム群において、

前記第三の記憶システムは、前記第三の記憶システム自身の処理負荷に応じて、前記所定の時間間隔を決定するように制御するものであることを特徴とする記憶システム群。

【請求項 3 4】

請求項 2 9 に記載の記憶システム群において、

前記第三の記憶システムは、前記第二の記憶システム内の前記第三の記憶領域が保持している記憶容量についての情報を、前記第二の記憶システムから読み出し、前記読み出された情報に応じて、前記所定の時間間隔を決定するように制御するものであることを特徴とする記憶システム群。

【請求項 3 5】

請求項 2 9 に記載の記憶システム群において、

前記第三の記憶領域に書き込まれる更新情報は、前記第二の記憶領域に書き込まれるデータの更新順序に関するものであることを特徴とする記憶システム群。

【請求項 3 6】

請求項 3 5 に記載の記憶システム群において、

前記第二の記憶システムは、前記第三の記憶領域についての管理情報を所有するものであり、

前記第三の記憶システムは、前記第二の記憶システムが所有している前記第三の記憶領域についての管理情報を、前記第二の記憶システムから読み出し、前記読み出された管理情報に応じて、前記所定の時間間隔を決定するように制御するものであることを特徴とする記憶システム群。

【請求項 3 7】

請求項 2 9 に記載の記憶システム群において、

前記第三の記憶システムは、前記第二の記憶領域に格納されたデータ及び前記データに関する更新情報の送付を要求するコマンドを、前記所定の時間間隔毎に、前記第二の記憶システムに対して送信するものであり、

前記第二の記憶システムは、前記コマンドに応じて、前記第三の記憶領域に格納されたデータ及び前記データに関する更新情報を前記第三の記憶システムに対して送信するものであることを特徴とする記憶システム群。

【請求項 3 8】

請求項 2 9 に記載の記憶システム群において、

前記第三の記憶システムは、前記第二の記憶システム内の前記第二の記憶領域

に対応する第五の記憶領域を有しており、前記第四の記憶領域に格納されているデータ及び更新情報を基にして、前記第五の記憶領域に対してデータを格納することを特徴とする記憶システム群。

【請求項 3 9】

請求項 3 8 に記載の記憶システム群において、

前記第三の記憶システムは、前記第三の記憶システム自身の処理負荷に応じて、前記第五の記憶領域に対するデータの格納処理のタイミングを制御することを特徴とする記憶システム群。

【発明の詳細な説明】

【 0 0 0 1 】

【発明の属する技術分野】

本発明は記憶システムに関し、特に複数の記憶システム間でのデータの複製に関する。

【 0 0 0 2 】

【従来の技術】

近年、常に顧客に対して継続したサービスを提供するために、第一の記憶システムに障害が発生した場合でもデータ処理システムがサービスを提供できるよう、記憶システム間でのデータの複製に関する技術が重要になっている。第一の記憶システムに格納された情報を第二および第三の記憶システムに複製する技術として、以下の特許文献に開示された技術が存在する。

【 0 0 0 3 】

米国特許 5 1 7 0 4 8 0 号公報には、第一の記憶システムに接続された第一の計算機が、第一の記憶システムに格納されたデータを、第一の計算機と第二の計算機間の通信リンクを介し、第二の計算機に転送し、第二の計算機が第二の計算機に接続された第二の記憶システムに転送する技術が開示されている。

【 0 0 0 4 】

米国特許 6 2 0 9 0 0 2 号公報には、第一の記憶システムが、第一の記憶システムに格納されたデータを、第二の記憶システムに転送し、さらに、第二の記憶システムが、第三の記憶システムに転送する技術が開示されている。計算機と第

一の記憶システムとは通信リンクにより接続され、第一の記憶システムと第二の記憶システムとは通信リンクにより接続され、さらに、第二の記憶システムと第三の記憶システムとは通信リンクにより接続されている。第一の記憶システムは、複製対象の第一の論理ボリュームを保持する。第二の記憶システムは、第一の論理ボリュームの複製である第二の論理ボリューム、及び、第二の論理ボリュームの複製である第三の論理ボリュームを保持する。第三の記憶システムは、第三の論理ボリュームの複製である第四の論理ボリュームを保持する。この特許文献において、第二の記憶システムは、第二の論理ボリュームから第三の論理ボリュームへのデータ複製処理と、第三の論理ボリュームから第四の論理ボリュームへのデータ複製処理とを排他的に実行する。

【0 0 0 5】

【特許文献 1】 米国特許 5 1 7 0 4 8 0 号公報

【特許文献 2】 米国特許 6 2 0 9 0 0 2 号公報

【発明が解決しようとする課題】

米国特許 5 1 7 0 4 8 0 号公報に開示された技術は、データの複製のために、第一の計算機および第二の計算機を常に使用する。第一の計算機は、通常業務を行っており、第一の計算機にかかるデータ複製処理の負荷は無視できない。さらに、複製のためのデータは、第一の計算機と第一の記憶システム間の通信リンクを使用するため、通常業務のために必要なデータ転送と衝突し、通常業務に必要なデータ参照、データ更新時間が長くなるという課題がある。

【0 0 0 6】

米国特許 6 2 0 9 0 0 2 号公報に開示された技術は、複製を行うデータ量の倍の記憶容量が第二の記憶システム及び第三の記憶システムに必要となる。また、複製対象のデータ量が多いことにより、データ複製処理に費やされる時間が長く、第三の記憶システムのデータは古いものになってしまう。その結果、第三の記憶システムのデータを用いて業務が再開される場合、第三の記憶システムのデータを最新とするまでの時間が長くなり、業務再開までの時間が延びるという課題がある。さらに、かかる文献において、第一の記憶システムは、第一の記憶システム内のデータ更新処理に加えて第二の記憶システムとの間でのデータ行進処理

が終了した時点で、上位の計算機にデータ更新完了報告を行う。したがって、計算機からのデータ更新に費やされる時間が長く、第一の記憶システムと第二の記憶システムとの間の距離が遠くなればなるほど、データ更新に費やされる時間は長くなる。その結果、かかる文献に開示された技術によれば、各記憶システム間の距離をあまり遠くすることができないという課題もある。

【0 0 0 7】

本発明の目的は、記憶システムの上位の計算機に影響を与えず、複数の記憶システム間でデータ転送又はデータの複製をするものである。さらに、記憶システムと計算機との間の通信にも影響を与えないものである。

【0 0 0 8】

さらに、本発明の目的は、複数の記憶システム内に保持するデータ格納領域を少なくすることができるものである。さらに、複数の記憶システムの上位の計算機の業務に影響を与えることのないように、高速かつ効率的に複数の記憶システム間でデータ転送又はデータの複製をするものである。

【0 0 0 9】

【課題を解決するための手段】

これらの課題を解決するために、本発明において、第一の記憶システムは、第一の記憶システムに格納されたデータの更新に関する情報をジャーナルとして格納する。ジャーナルは、具体的には、更新に用いられたデータのコピーと更新時のライト命令等の更新情報とによって構成される。さらに、第二の記憶システムは、第一の記憶システムと第二の記憶システム間の通信線を介して、前記ジャーナルを取得する。第二の記憶システムは、第一の記憶システムが保持するデータの複製を保持しており、前記ジャーナルを用いて、第一の記憶システムでのデータ更新順に、第一の記憶システムのデータと対応するデータを更新する。

【0 0 1 0】

本発明においては、第二の記憶システムは、第一の記憶システムが格納する第一の記憶領域に格納されたデータの複製を格納する第二の記憶領域を保持し、第二の記憶領域のデータの更新に関する情報をジャーナルとしてジャーナル専用の第三の記憶領域に格納する。第三の記憶領域の記憶容量は、第二の記憶領域より

少ない記憶容量とすることが可能である。さらに、第三の記憶システムは、第二の記憶システムと第三の記憶システム間の通信線を介して、前記ジャーナルを取得し、ジャーナル専用の第四の記憶領域に格納する。第四の記憶領域の記憶容量は、第二の記憶領域より少ない記憶容量とすることが可能である。第三の記憶システムは、第二の記憶領域に格納されたデータの複製を格納する第五の記憶領域を保持し、前記ジャーナルを用いて、第二の記憶領域でのデータ更新順に、第二の記憶領域に対応する第五の記憶領域のデータを更新する。

【0011】

【発明の実施の形態】

本発明によるデータ処理システムの実施形態を図面により詳細に説明する。

【0012】

図1は、本発明の一実施形態の論理的な構成を示すブロック図である。

【0013】

本発明の一実施形態は、ホストコンピュータ180と記憶システム100Aを接続バス190により接続し、記憶システム100Aと記憶システム100Aに保存されたデータの複製を保持する記憶システム100Bを接続バス200により接続した構成である。以下の説明において、複製対象のデータを保持する記憶システム100と複製データを保持する記憶システム100との区別を容易とするために、複製対象のデータを保持する記憶システム100を正記憶システム100A、複製データを保持する記憶システム100を副記憶システム100Bとよぶこととする。なお、記憶システムの記憶領域は、分割して管理されており、分割された記憶領域を論理ボリュームと呼ぶこととする。

【0014】

記憶システム100の記憶領域は、分割して管理されており、分割した記憶領域を論理ボリューム230とよぶこととする。論理ボリューム230の容量および記憶システム100内の物理的な格納位置（物理アドレス）は、記憶システム100に接続したコンピュータ等の保守端末もしくはホストコンピュータ180を用いて指定できる。各論理ボリューム230の物理アドレスは、後述するボリューム情報400に保存する。物理アドレスは、例えば、記憶システム100内

の記憶装置 1 5 0 を識別する番号（記憶装置番号）と記憶装置内の記憶領域を一意に示す数値、例えば、記憶装置の記憶領域の先頭からの位置である。以下の説明では、物理アドレスは、記憶装置番号と記憶装置の記憶領域の先頭からの位置の組とする。以下の説明では、論理ボリュームは、1 つの記憶装置の記憶領域であるが、論理アドレスと物理アドレスの変換により、1 つの論理ボリュームを複数の記憶装置の記憶領域に対応づけることも可能である。

【0 0 1 5】

記憶システム 1 0 0 が保存しているデータの参照、更新は、論理ボリュームを識別する番号（論理ボリューム番号）と記憶領域を一意に示す数値、例えば、論理ボリュームの記憶領域の先頭からの位置により一意に指定することができ、以下、論理ボリューム番号と論理ボリュームの記憶領域の先頭からの位置（論理アドレス内位置）の組を論理アドレスとよぶ。

【0 0 1 6】

以下の説明において、複製対象のデータと複製データとの区別を容易とするために、複製対象の論理ボリューム 2 3 0 を正論理ボリューム、複製データである論理ボリューム 2 3 0 を副論理ボリュームとよぶこととする。一对の正論理ボリュームと副論理ボリュームをペアとよぶ。正論理ボリュームと副論理ボリュームの関係および状態等は後述するペア情報 5 0 0 に保存する。

【0 0 1 7】

論理ボリューム間のデータの更新順序を守るために、グループという管理単位を設ける。例えば、ホストコンピュータ 1 8 0 が、正論理ボリューム 1 のデータ 1 を更新し、その後、データ 1 を読み出し、データ 1 の数値を用いて、正論理ボリューム 2 のデータ 2 を更新する処理を行うとする。正論理ボリューム 1 から副論理ボリューム 1 へのデータ複製処理と、正論理ボリューム 2 から副論理ボリューム 2 へのデータ複製処理とが独立に行われる場合、副論理ボリューム 1 へのデータ 1 の複製処理より前に、副論理ボリューム 2 へのデータ 2 の複製処理が行われる場合がある。副論理ボリューム 2 へのデータ 2 の複製処理と副論理ボリューム 1 へのデータ 1 の複製処理との間に、故障等により副論理ボリューム 1 へのデータ 1 の複製処理が停止した場合、副論理ボリューム 1 と副論理ボリューム 2 の

データの整合性がなくなる。このような場合にも副論理ボリューム1と副論理ボリューム2のデータの整合性を保つために、データの更新順序を守る必要のある論理ボリュームは、同じグループに登録し、データの更新毎に、後述するグループ情報600の更新番号を割り当て、更新番号順に、副論理ボリュームに複製処理を行う。例えば、図1では、正記憶システム100Aの論理ボリューム（DATA1）と論理ボリューム（DATA2）がグループ1を構成する。論理ボリューム（DATA1）の副製である論理ボリューム（COPY1）と論理ボリューム（DATA2）の複製である論理ボリューム（COPY2）は、副記憶システム内でグループ1を構成する。

【0018】

データ複製対象である正論理ボリュームのデータを更新する場合、副論理ボリュームのデータを更新するために、後述するジャーナルを作成し、正記憶システム100A内の論理ボリュームに保存する。本実施例の説明では、グループ毎にジャーナルのみを保存する論理ボリューム（以下、ジャーナル論理ボリュームとよぶ）を割り当てる。図1では、グループ1に論理ボリューム（JNL1）を割り当てている。

【0019】

副記憶システム100Bのグループにもジャーナル論理ボリュームを割り当てる。ジャーナル論理ボリュームは、正記憶システム100Aから副記憶システム100Bに転送したジャーナルを保存するために使用する。ジャーナルをジャーナル論理ボリュームに保存することにより、例えば、副記憶システム100Bの負荷が高い場合、ジャーナル受信時に副論理ボリュームのデータ更新を行わず、暫く後、副記憶システム100Bの負荷が低い時に、副論理ボリュームのデータを更新することもできる。さらに、接続線200が複数ある場合、正記憶システム100Aから副記憶システム100Bへのジャーナルの転送を多重に行い、接続線200の転送能力を有効に利用することができる。更新順番のため、副記憶システム100Bに多くのジャーナルが溜まる可能性があるが、副論理ボリュームのデータ更新に直ぐに使用できないジャーナルは、ジャーナル論理ボリュームに退避することにより、キャッシュメモリを開放することができる。図1では、

副記憶システム内のグループ 1 に論理ボリューム（JNL 2）を割り当てている。

【0020】

ジャーナルは、ライトデータと更新情報とから構成する。更新情報は、ライトデータを管理するための情報で、ライト命令を受信した時刻、グループ番号、後述するグループ情報 6 0 0 の更新番号、ライト命令の論理アドレス、ライトデータのデータサイズ、ライトデータを格納したジャーナル論理ボリュームの論理アドレス等からなる。更新情報は、ライト命令を受信した時刻と更新番号のどちらか一方のみを保持してもよい。ホストコンピュータ 1 8 0 からのライト命令内にライト命令の作成時刻が存在する場合は、ライト命令を受信した時刻の代わりに、当該ライト命令内の作成時刻を使用してもよい。図 3 と図 2 1 を用いて、ジャーナルの更新情報の例を説明する。更新情報 3 1 0 は、1 9 9 9 年 3 月 1 7 日の 2 2 時 2 0 分 1 0 秒に受信したライト命令を記憶する。当該ライト命令は、ライトデータを論理ボリューム番号 1 の記憶領域の先頭から 7 0 0 の位置に格納する命令であり、データサイズは 3 0 0 である。ジャーナルのライトデータは、論理ボリューム番号 4（ジャーナル論理ボリューム）の記憶領域の先頭から 1 5 0 0 の位置から格納されている。論理ボリューム番号 1 の論理ボリュームはグループ 1 に属し、グループ 1 のデータ複製開始から 4 番目のデータ更新であることがわかる。

【0021】

ジャーナル論理ボリュームは、例えば、図 3 に示すように、更新情報を格納する記憶領域（更新情報領域）とライトデータを格納する記憶領域（ライトデータ領域）に分割して使用する。更新情報領域は、更新情報領域の先頭から、更新番号の順に格納し、更新情報領域の終端に達すると、更新情報領域の先頭から格納する。ライトデータ領域は、ライトデータ領域の先頭からライトデータを格納し、ライトデータ領域の終端に達すると、ライトデータ領域の先頭から格納する。更新情報領域およびライトデータ領域の比は、固定値でもよいし、保守端末あるいはホストコンピュータ 1 8 0 により設定可能としてもよい。これらの情報は、後述するポインタ情報 7 0 0 内に保持する。以下の説明では、ジャーナル論理ボ

リュームを更新情報とライトデータの領域に分割して、ジャーナル論理ボリュームを使用するが、論理ボリュームの先頭から、ジャーナル、つまり、更新情報とライトデータを連続に格納する方式を採用してもよい。

【0022】

図1を用いて、記憶システム100Aの正論理ボリュームへのデータ更新を副記憶システム100Bの副論理ボリュームに反映する動作について概説する。

【0023】

(1) 記憶システム100Aは、ホストコンピュータ180から正論理ボリューム内のデータに対するライト命令を受信すると、後述する命令受信処理210およびリードライト処理220によって、正論理ボリューム(DATA1)内のデータ更新と、ジャーナル論理ボリューム(JNL1)にジャーナルの保存を行う(図1の270)。

【0024】

(2) 記憶システム100Bは、後述するジャーナルリード処理240によって、記憶システム100Aからジャーナルをリードし、リードライト処理220によって、ジャーナル論理ボリューム(JNL2)にジャーナルを保存する(図1の280)。

【0025】

(3) 記憶システム100Aは、記憶システム100Bからジャーナルをリードする命令を受信すると、後述する命令受信処理210およびリードライト処理220によって、ジャーナル論理ボリューム(JNL1)からジャーナルを読み出し、記憶システム100Bに送信する(図1の280)。

【0026】

(4) 記憶システム100Bは、後述するリストア処理250およびリードライト処理220によって、ポインタ情報700を用いて、更新番号の昇順に、ジャーナル論理ボリューム(JNL2)からジャーナルを読み出し、副論理ボリューム(COPY1)のデータを更新する(図1の290)。

【0027】

記憶システム100の内部構造を図2に示す。記憶システム100は、1つ以

上のホストアダプタ 110、1つ以上のディスクアダプタ 120、1つ以上のキャッシュメモリ 130、1つ以上の共有メモリ 140、1つ以上の記憶装置 150、1つ以上のコモンバス 160、1つ以上の接続線 170を備えて構成される。ホストアダプタ 110、ディスクアダプタ 120、キャッシュメモリ 130、共有メモリ 140はコモンバス 160により相互間が接続されている。コモンバス 160は、コモンバス 160の障害時のために2重化されてもよい。ディスクアダプタ 120と記憶装置 150とは接続線 170によって接続されている。また、図示していないが、記憶システム 100の設定、監視、保守等を行うための保守端末が全てのホストアダプタ 110とディスクアダプタ 120とに専用線を用いて接続されている。

【0028】

ホストアダプタ 110は、ホストコンピュータ 180とキャッシュメモリ 130間のデータ転送を制御する。ホストアダプタ 110は、接続線 190および接続線 200によりホストコンピュータ 180もしくは、他の記憶システム 100と接続される。ディスクアダプタ 120は、キャッシュメモリ 130と記憶装置 150との間のデータ転送を制御する。キャッシュメモリ 130は、ホストコンピュータ 180から受信したデータあるいは記憶装置 150から読み出したデータを一時的に保持するメモリである。共有メモリ 140は、記憶システム 100内の全てのホストアダプタ 110とディスクアダプタ 120とが共有するメモリである。

【0029】

ボリューム情報 400は、論理ボリュームを管理する情報であり、ボリューム状態、フォーマット形式、容量、ペア番号、物理アドレスを保持する。図4にボリューム情報 400の一例を示す。ボリューム情報 400は、ホストアダプタ 110およびディスクアダプタ 120から参照可能なメモリ、例えば管理メモリ 140に保存される。ボリューム状態は、“正常”、“正”、“副”、“異常”、“未使用”のいずれかを保持する。ボリューム状態が“正常”もしくは“正”である論理ボリューム 230は、ホストコンピュータ 180から正常にアクセス可能な論理ボリューム 230であることを示す。ボリューム状態が“副”である論

理ボリューム 2 3 0 は、ホストコンピュータ 1 8 0 からのアクセスを許可してもよい。ボリューム状態が“正”である論理ボリューム 2 3 0 は、データの複製が行われている論理ボリューム 2 3 0 であることを示す。ボリューム状態が“副”である論理ボリューム 2 3 0 は、複製に使用されている論理ボリューム 2 3 0 であることを示す。ボリューム状態が“異常”の論理ボリューム 2 3 0 は、障害により正常にアクセスできない論理ボリューム 2 3 0 であることを示す。障害とは、例えば、論理ボリューム 2 3 0 を保持する記憶装置 1 5 0 の故障である。ボリューム状態が“未使用”の論理ボリューム 2 3 0 は、使用していない論理ボリューム 2 3 0 であることを示す。ペア番号は、ボリューム状態が“正”もしくは“副”の場合に有効であり、後述するペア情報 5 0 0 を特定するためのペア番号を保持する。図 4 に示す例では、論理ボリューム 1 は、フォーマット形式が O P E N 3、容量が 3 G B、記憶装置番号 1 の記憶装置 1 5 0 の記憶領域の先頭からデータが格納されており、アクセス可能であり、データの複製対象であることを示す。

【 0 0 3 0 】

ペア情報 5 0 0 は、ペアを管理する情報であり、ペア状態、正記憶システム番号、正論理ボリューム番号、副記憶システム番号、副論理ボリューム番号、グループ番号、コピー済みアドレスを保持する。図 5 にペア情報 5 0 0 の一例を示す。ペア情報 5 0 0 は、ホストアダプタ 1 1 0 およびディスクアダプタ 1 2 0 から参照可能なメモリ、例えば管理メモリ 1 4 0 に保存する。ペア状態は、“正常”、“異常”、“未使用”、“コピー未”、“コピー中”のいずれかを保持する。ペア状態が“正常”の場合は、正論理ボリューム 2 3 0 のデータ複製が正常に行われていることを示す。ペア状態が“異常”の場合は、障害により正論理ボリューム 2 3 0 の複製が行えないことを示す。障害とは、例えば、接続バス 2 0 0 の断線などである。ペア状態が“未使用”の場合は、当該ペア番号の情報は有効でないことを示す。ペア状態が“コピー中”の場合は、後述する初期コピー処理中であることを示す。ペア状態が“コピー未”の場合は、後述する初期コピー処理が未だ行われていないことを示す。正記憶システム番号は、正論理ボリューム 2 3 0 を保持する正記憶システム 1 0 0 A を特定する番号を保持する。副記憶シス

テム番号は、副論理ボリューム 2 3 0 を保持する副記憶システム 1 0 0 B を特定する番号を保持する。グループ番号は、正記憶システムの場合は、正論理ボリュームが属するグループ番号を保持する。副記憶システムの場合は、副論理ボリュームが属するグループ番号を保持する。コピー済みアドレスは、後述する初期コピー処理にて説明する。図 5 のペア情報 1 は、データ複製対象が正記憶システム 1 の正論理ボリューム 1 であり、データ複製先が副記憶システム 2 の副論理ボリューム 1 であり、正常にデータ複製処理が行われていることを示す。

【 0 0 3 1 】

グループ情報 6 0 0 は、グループ状態、ペア集合、ジャーナル論理ボリューム番号、更新番号を保持する。図 6 にグループ情報 6 0 0 の一例を示す。グループ情報 6 0 0 は、ホストアダプタ 1 1 0 およびディスクアダプタ 1 2 0 から参照可能なメモリ、例えば管理メモリ 1 4 0 に保存する。グループ状態は、“正常”、“異常”、“未使用”のいずれかを保持する。グループ状態が“正常”の場合は、ペア集合の少なくともひとつのペア状態が“正常”であることを示す。グループ状態が“異常”の場合は、ペア集合の全てのペア状態が“異常”であることを示す。グループ状態が“未使用”の場合は、当該グループ番号の情報は有効でないことを示す。ペア集合は、正記憶システムの場合は、グループ番号が示すグループに属する全ての正論理ボリュームのペア番号を保持する。副記憶システムの場合は、グループ番号が示すグループに属する全ての副論理ボリュームのペア番号を保持する。ジャーナル論理ボリューム番号は、当該グループ番号のグループに属するジャーナル論理ボリューム番号を示す。更新番号は、初期値は 1 であり、グループ内の正論理ボリュームに対しデータの書き込みが行われると、1 を加える。更新番号は、ジャーナルの更新情報に記憶し、副記憶システム 1 0 0 B にて、データの更新順を守るために使用する。例えば、図 6 のグループ情報 1 は、ペア情報 1, 2 から、正論理ボリューム 1, 2 と、ジャーナル論理ボリューム 4 から構成されており、正常にデータの複製処理が行われていることを示す。

【 0 0 3 2 】

ポインタ情報 7 0 0 は、グループ毎に保持し、当該グループのジャーナル論理ボリュームを管理する情報であり、更新情報領域先頭アドレス、ライトデータ領

域先頭アドレス、更新情報最新アドレス、更新情報最古アドレス、ライトデータ最新アドレス、ライトデータ最古アドレス、リード開始アドレス、リトライ開始アドレスを保持する。図 7 および図 8 にポインタ情報 7 0 0 の一例を示す。更新情報領域先頭アドレスは、ジャーナル論理ボリュームの更新情報を格納する記憶領域（更新情報領域）の先頭の論理アドレスを保持する。ライトデータ領域先頭アドレスは、ジャーナル論理ボリュームのライトデータを格納する記憶領域（ライトデータ領域）の先頭の論理アドレスを保持する。更新情報最新アドレスは、次にジャーナルを格納する場合に、更新情報の保存に使用する先頭の論理アドレスを保持する。更新情報最古アドレスは、最古の（更新番号が小さい）ジャーナルの更新情報を保存する先頭の論理アドレスを保持する。ライトデータ最新アドレスは、次にジャーナルを格納する場合に、ライトデータの保存に使用する先頭の論理アドレスを保持する。ライトデータ最古アドレスは、最古の（更新番号が小さい）ジャーナルのライトデータを保存する先頭の論理アドレスを保持する。リード開始アドレスとリトライ開始アドレスは、正記憶システム 1 0 0 A のみで使用し、後述するジャーナルリード受信処理にて使用する。図 7 および図 8 のポインタ情報 7 0 0 の例では、ジャーナルの管理情報を保存する領域（更新情報領域）は、論理ボリューム 4 の記憶領域の先頭から 6 9 9 の位置までであり、ジャーナルのライトデータを保存する領域（ライトデータ領域）は、論理ボリューム 4 の記憶領域の 7 0 0 の位置から 2 6 9 9 の位置までである。ジャーナルの管理情報は、論理ボリューム 4 の記憶領域の 2 0 0 の位置から 4 9 9 の位置まで保存されており、次のジャーナルの管理情報は、論理ボリューム 4 の記憶領域の 5 0 0 の位置から保存する。ジャーナルのライトデータは論理ボリューム 4 の記憶領域の 1 3 0 0 の位置から 2 1 9 9 の位置まで保存されており、次のジャーナルのライトデータは、論理ボリューム 4 の記憶領域の 2 2 0 0 の位置から保存する。

【 0 0 3 3 】

下記の説明では、1つのグループに1つのジャーナル論理ボリュームを割り当てる形態で説明しているが、1つのグループに複数のジャーナル論理ボリュームを割り当ててもよい。例えば、1つのグループに2つのジャーナル論理ボリュームを割り当て、ジャーナル論理ボリューム毎にポインタ情報 7 0 0 を設け、交互

にジャーナルを格納する。これにより、ジャーナルの記憶装置 1 5 0 への書き込みが分散でき、性能の向上が見込める。さらに、ジャーナルのリード性能も向上する。別の例としては、1つのグループに2つのジャーナル論理ボリュームを割り当て、通常は、1つのジャーナル論理ボリュームのみを使用する。もう一方のジャーナル論理ボリュームは、使用しているジャーナル論理ボリュームの性能が低下した場合に使用する。性能が低下する例は、ジャーナル論理ボリュームが、複数の記憶装置 1 5 0 から構成され、R A I D 5 の方式でデータを保持しており、構成する記憶装置 1 5 0 の一台が故障中の場合である。

【 0 0 3 4 】

なお、上述のボリューム情報 4 0 0、ペア情報 5 0 0、グループ情報 6 0 0、及びポインタ情報 7 0 0 等は、共有メモリ 1 4 0 に格納されていることが好ましい。しかし、本実施例は、この場合に限られず、これらの情報を、キャッシュメモリ 1 3 0、ホストアダプタ 1 1 0、ディスクアダプタ 1 2 0、その他記憶装置 1 5 0 に集中して格納または分散して格納することもよい。

【 0 0 3 5 】

次に、正記憶システム 1 0 0 A から副記憶システム 1 0 0 B に対して、データ複製を開始する手順を図 9、図 1 0 を用いて説明する。

【 0 0 3 6 】

(1) グループ作成について説明する (ステップ 9 0 0)。ユーザは、保守端末あるいはホストコンピュータ 1 8 0 を使用して、正記憶システム 1 0 0 A のグループ情報 6 0 0 を参照し、グループ状態が“未使用”のグループ番号 A を取得する。ユーザは、保守端末あるいはホストコンピュータ 1 8 0 を使用して、グループ番号 A を指定し、グループ作成指示を正記憶システム 1 0 0 A に行う。

【 0 0 3 7 】

グループ作成指示を受けて、正記憶システム 1 0 0 A は、指定されたグループ番号 A のグループ状態を“正常”に変更する。

【 0 0 3 8 】

同様に、ユーザは、副記憶システム 1 0 0 B のグループ情報 6 0 0 を参照し、グループ状態が“未使用”のグループ番号 B を取得する。ユーザは、保守端末あ

るいはホストコンピュータ 1 8 0 を使用して、副記憶システム 1 0 0 B とグループ番号 B を指定し、グループ作成指示を正記憶システム 1 0 0 A に行う。正記憶システム 1 0 0 A は、受信したグループ作成指示を副記憶システム 1 0 0 B に転送する。副記憶システム 1 0 0 B は、指定されたグループ番号 B のグループ状態を“正常”に変更する。

【0 0 3 9】

ユーザは、副記憶システム 1 0 0 B の保守端末あるいは副記憶システム 1 0 0 B に接続したホストコンピュータ 1 8 0 を使用して、グループ番号 B を指定し、グループ作成指示を副記憶システム 1 0 0 B に行ってもよい。

【0 0 4 0】

(2) ペア登録について説明する (ステップ 9 1 0)。ユーザは、保守端末あるいはホストコンピュータ 1 8 0 を使用して、データ複製対象を示す情報とデータ複製先を示す情報を指定し、ペア登録指示を正記憶システム 1 0 0 A に行う。データ複製対象を示す情報は、データ複製対象のグループ番号 A と正論理ボリューム番号である。データ複製先を示す情報は、複製データを保存する副記憶システム 1 0 0 B とグループ番号 B、副論理ボリューム番号である。

【0 0 4 1】

前記ペア登録指示を受けて、正記憶システム 1 0 0 A は、ペア情報 5 0 0 からペア情報が“未使用”のペア番号を取得し、ペア状態を“コピー未”に、正記憶システム番号に、正記憶システム 1 0 0 A を示す正記憶システム番号を、正論理ボリューム番号に、指示された正論理ボリューム番号を、副記憶システム番号に、指示された副記憶システム番号を、副論理ボリューム番号に、指示された副論理ボリューム番号を、グループ番号に、指示されたグループ番号 A を設定する。正記憶システム 1 0 0 A は、指示されたグループ番号 A のグループ情報 6 0 0 のペア集合に前記取得したペア番号を追加し、正論理ボリューム番号のボリューム状態を“正”に変更する。

【0 0 4 2】

正記憶システム 1 0 0 A は、正記憶システム 1 0 0 A を示す正記憶システム番号、ユーザから指定されたグループ番号 B、正論理ボリューム番号、および副論

理ボリューム番号を副記憶システム 1 0 0 B に指示する。副記憶システム 1 0 0 B は、ペア情報 5 0 0 から未使用のペア番号を取得し、ペア状態を “コピー未” に、正記憶システム番号に、記憶システム 1 0 0 A を示す正記憶システム番号を、正論理ボリューム番号に、指示された正論理ボリューム番号を、副記憶システム番号に、副記憶システム B を示す副記憶システム番号を、副論理ボリューム番号に、指示された副論理ボリューム番号を、グループ番号に、指示されたグループ番号 B を設定する。

【 0 0 4 3 】

副記憶システム 1 0 0 B は、指示されたグループ番号 B のグループ情報 6 0 0 のペア集合に前記取得したペア番号を追加し、副論理ボリューム番号のボリューム状態を “副” に変更する。

【 0 0 4 4 】

以上の動作を全てのデータ複製対象のペアに対して行う。

【 0 0 4 5 】

前記の説明では、論理ボリュームのグループへの登録と、論理ボリュームのペアの設定を同時に行う処理を説明したが、それぞれ個別に行ってもよい。

【 0 0 4 6 】

(3) ジャーナル論理ボリューム登録について説明する (ステップ 9 2 0) 。ユーザは、保守端末あるいはホストコンピュータ 1 8 0 を使用して、ジャーナルの保存に使用する論理ボリューム (ジャーナル論理ボリューム) をグループに登録する指示 (ジャーナル論理ボリューム登録指示) を正記憶システム 1 0 0 A に行う。ジャーナル論理ボリューム登録指示は、グループ番号と論理ボリューム番号からなる。

【 0 0 4 7 】

正記憶システム 1 0 0 A は、指示されたグループ番号のグループ情報 6 0 0 のジャーナル論理ボリューム番号に指示された論理ボリューム番号を登録する。当該論理ボリュームのボリューム情報 4 0 0 のボリューム状態を “正常” に設定する。

【 0 0 4 8 】

同様に、ユーザは、保守端末あるいはホストコンピュータ 1 8 0 を使用して、副記憶システム 1 0 0 B のボリューム情報 4 0 0 を参照し、副記憶システム 1 0 0 B、グループ番号 B、ジャーナル論理ボリュームとして使用する論理ボリューム番号を指定し、ジャーナル論理ボリューム登録を正記憶システム 1 0 0 A に行う。正記憶システム 1 0 0 A は、ジャーナル論理ボリューム登録指示を副記憶システム 1 0 0 B に転送する。副記憶システム 1 0 0 B は、指示されたグループ番号 B のグループ情報 6 0 0 のジャーナル論理ボリューム番号に指示された論理ボリューム番号を登録する。当該論理ボリュームのボリューム情報 4 0 0 のボリューム状態を“正常”に設定する。

【 0 0 4 9 】

ユーザは、副記憶システム 1 0 0 B の保守端末あるいは副記憶システム 1 0 0 B に接続したホストコンピュータ 1 8 0 を使用して、グループ番号、ジャーナル論理ボリュームとして使用する論理ボリューム番号を指定し、ジャーナル論理ボリューム登録指示を副記憶システム 1 0 0 B に行ってもよい。

【 0 0 5 0 】

以上の動作を全てのジャーナル論理ボリュームとして使用する論理ボリュームに対して行う。ステップ 9 1 0 とステップ 9 2 0 の順は不同ではない。

【 0 0 5 1 】

(4) データ複製処理開始について説明する (ステップ 9 3 0) 。ユーザは、保守端末あるいはホストコンピュータ 1 8 0 を使用して、データ複製処理を開始するグループ番号を指定し、データ複製処理の開始を正記憶システム 1 0 0 A に指示する。正記憶システム 1 0 0 A は、指示されたグループに属する全てのペア情報 4 0 0 のコピー済みアドレスを 0 に設定する。

【 0 0 5 2 】

正記憶システム 1 0 0 A は、副記憶システム 1 0 0 B に、後述するジャーナルリード処理およびリストア処理の開始を指示する。

【 0 0 5 3 】

正記憶システム 1 0 0 A は、後述する初期コピー処理を開始する。

【 0 0 5 4 】

(5) 初期コピー終了について説明する (ステップ 9 4 0)。

【0 0 5 5】

初期コピーが終了すると、正記憶システム 1 0 0 A は、初期コピー処理の終了を副記憶システム 1 0 0 B に通知する。副記憶システム 1 0 0 B は、指示されたグループに属する全ての副論理ボリュームのペア状態を“正常”に変更する。

【0 0 5 6】

図 1 0 は、初期コピー処理のフローチャートである。初期コピー処理は、データ複製対象の正論理ボリュームの全記憶領域に対し、ペア情報 5 0 0 のコピー済みアドレスを用い、記憶領域の先頭から順に、単位サイズ毎にジャーナルを作成する。コピー済みアドレスは、初期値は 0 であり、ジャーナルの作成毎に、作成したデータ量を加算する。論理ボリュームの記憶領域の先頭から、コピー済みアドレスの一つ前までは、初期コピー処理にてジャーナルは作成済みである。初期コピー処理を行うことにより、正論理ボリュームの更新されていないデータを副論理ボリュームに転送することが可能となる。以下の説明では、正記憶システム 1 0 0 A 内のホストアダプタ A が処理を行うように記載しているが、ディスクアダプタ 1 2 0 が行ってもよい。

【0 0 5 7】

(1) 正記憶システム 1 0 0 A 内のホストアダプタ A は、処理対象のグループに属するペアでペア状態が“コピー未”である正論理ボリューム A を得、ペアの状態を“コピー中”に変更し、以下の処理を繰り返す (ステップ 1 0 1 0、1 0 2 0)。正論理ボリューム A が存在しない場合は、処理を終了する (ステップ 1 0 3 0)。

【0 0 5 8】

(2) ステップ 1 0 2 0 にて、論理ボリューム A が存在した場合、ホストアダプタ A は、単位サイズ (例えば、1 MB) のデータを対象にジャーナルを作成する。ジャーナル作成処理は後述する (ステップ 1 0 4 0)。

【0 0 5 9】

(3) ホストアダプタ A は、コピー済みアドレスに作成したジャーナルのデータサイズを加算する (ステップ 1 0 5 0)。

【0060】

(4) コピー済みアドレスが、正論理ボリュームAの容量に達するまで、上記処理を繰り返す(ステップ1060)。コピー済みアドレスが、正論理ボリュームAの容量と等しくなった場合、正論理ボリュームAの全記憶領域に対してジャーナルを作成したため、ペア状態を“正常”に更新し、他の正論理ボリュームの処理を開始する(ステップ1070)。

【0061】

前記のフローチャートでは、論理ボリュームを1つずつ対象とするように説明したが、複数の論理ボリュームを同時に処理してもよい。

【0062】

図11は命令受信処理210の処理を説明する図、図12は命令受信処理210のフローチャート、図13はジャーナル作成処理のフローチャートである。以下、これらを用いて、正記憶システム100Aが、ホストコンピュータ180からデータ複製対象の論理ボリューム230にライト命令を受信した場合の動作について説明する。

【0063】

(1) 記憶システム100A内のホストアダプタAは、ホストコンピュータからアクセス命令を受信する。アクセス命令は、リード、ライト、後述するジャーナルリード等の命令、命令対象の論理アドレス、データ量等を含んでいる。以下、アクセス命令内の論理アドレスを論理アドレスA、論理ボリューム番号を論理ボリュームA、論理ボリューム内位置を論理ボリューム内位置A、データ量をデータ量Aとする(ステップ1200)。

【0064】

(2) ホストアダプタAは、アクセス命令を調べる(ステップ1210、1215)。ステップ1215の調べで、アクセス命令がジャーナルリード命令の場合は、後述するジャーナルリード受信処理を行う(ステップ1220)。アクセス命令がジャーナルリード命令およびライト命令以外、例えば、リード命令の場合は、従来技術と同じようにリード処理を行う(ステップ1230)。

【0065】

(3) ステップ 1 2 1 0 の調べで、アクセス命令がライト命令の場合は、論理ボリューム A のボリューム情報 4 0 0 を参照し、ボリューム状態を調べる（ステップ 1 2 4 0）。ステップ 1 2 4 0 の調べで、論理ボリューム A のボリューム状態が、“正常” もしくは“正” 以外の場合は、論理ボリューム A へのアクセスは不可能なため、ホストコンピュータ 1 8 0 に異常終了を報告する（ステップ 1 2 4 5）。

【0 0 6 6】

(4) ステップ 1 2 4 0 の調べで、論理ボリューム A のボリューム状態が、“正常”、“正” のいずれかの場合は、ホストアダプタ A は、キャッシュメモリ 1 3 0 を確保し、ホストコンピュータ 1 8 0 にデータ受信の準備ができたことを通知する。ホストコンピュータ 1 8 0 は、その通知を受け、ライトデータを正記憶システム 1 0 0 A に送信する。ホストアダプタ A は、ライトデータを受信し、当該キャッシュメモリ 1 3 0 に保存する（ステップ 1 2 5 0、図 1 1 の 1 1 0 0）。

【0 0 6 7】

(5) ホストアダプタ A は、論理ボリューム A のボリューム状態を参照し、論理ボリューム A がデータ複製対象かどうかを調べる（ステップ 1 2 6 0）。ステップ 1 2 6 0 の調べで、ボリューム状態が、“正” である場合は、論理ボリューム A がデータ複製対象であるため、後述するジャーナル作成処理を行う（ステップ 1 2 6 5）。

【0 0 6 8】

(6) ステップ 1 2 6 0 の調べで、ボリューム状態が、“正常” である場合、もしくはステップ 1 2 6 5 のジャーナル作成処理の終了後、ホストアダプタ A は、ディスクアダプタ 1 2 0 にライトデータを記憶装置 1 5 0 に書き込むことを命令し（図 1 1 の 1 1 4 0）、ホストコンピュータ 1 8 0 に終了報告する（ステップ 1 2 7 0、1 2 8 0）。その後、当該ディスクアダプタ 1 2 0 は、リードライト処理により、記憶装置 1 5 0 にライトデータを保存する（図 1 1 の 1 1 1 0）。

【0 0 6 9】

次に、ジャーナル作成処理について説明する。

【0070】

(1) ホストアダプタ A は、ジャーナル論理ボリュームのボリューム状態を調べる（ステップ 1310）。ステップ 1310 の調べで、ジャーナル論理ボリュームのボリューム状態が、“異常” の場合は、ジャーナル論理ボリュームにジャーナルの格納が不可能なため、グループ状態を“異常” に変更し、処理を終了する（ステップ 1315）。この場合、ジャーナル論理ボリュームを正常な論理ボリュームに変更する等を行う。

【0071】

(2) ステップ 1310 の調べで、ジャーナル論理ボリュームが正常である場合、ジャーナル作成処理を継続する。ジャーナル作成処理は、初期コピー処理内の処理であるか、命令受信処理内の処理であるかによって処理が異なる（ステップ 1320）。ジャーナル作成処理が命令受信処理内の処理の場合は、ステップ 1330 からの処理を行う。ジャーナル作成処理が初期コピー処理内の場合は、ステップ 1370 からの処理を行う。

【0072】

(3) ジャーナル作成処理が命令受信処理内の処理の場合、ホストアダプタ A は、ライト対象の論理アドレス A が、初期コピー処理の処理対象となったかを調べる（ステップ 1330）。論理ボリューム A のペア状態が“コピー未” の場合は、後に初期コピー処理にてジャーナル作成処理が行われるため、ジャーナルを作成せずに処理を終了する（ステップ 1335）。論理ボリューム A のペア状態が“コピー中” の場合は、コピー済みアドレスが論理アドレス内位置 A と等しいもしくは、小さい場合は、後に初期コピー処理にてジャーナル作成処理が行われるため、ジャーナルを作成せずに処理を終了する（ステップ 1335）。上記以外、つまり、論理ボリューム A のペア状態が“コピー中” かつコピー済みアドレスが論理アドレス内位置 A 以上の場合もしくは、論理ボリューム A のペア状態が“正常” の場合は、既に初期コピー処理が完了しているため、ジャーナル作成処理を継続する。

【0073】

(4) 次に、ホストアダプタ A は、ジャーナルがジャーナル論理ボリュームに格納可能であることを調べる。ポインタ情報 7 0 0 を用い、更新情報領域の未使用領域の有無を調べる (ステップ 1 3 4 0)。ポインタ情報 7 0 0 の更新情報最新アドレスと更新情報最古アドレスが等しい場合は、更新情報領域に未使用領域が存在しないため、ジャーナル作成失敗として処理を終了する (ステップ 1 3 9 0)。

【0 0 7 4】

ステップ 1 3 4 0 の調べで、更新情報領域に未使用領域が存在する場合は、ポインタ情報 7 0 0 を用い、ライトデータ領域にライトデータが格納できるかを調べる (ステップ 1 3 4 5)。ライトデータ最新アドレスとデータ量 A の和が、ライトデータ最古アドレスと等しいもしくは大きい場合は、ライトデータ領域に格納できないため、ジャーナル作成失敗として処理を終了する (ステップ 1 3 9 0)。

【0 0 7 5】

(5) ジャーナルが格納可能である場合、ホストアダプタ A は、更新番号と更新情報を格納する論理アドレスとライトデータを格納する論理アドレスを取得し、更新情報をキャッシュメモリ 1 3 0 内に作成する。更新番号は、対象グループのグループ情報 6 0 0 から取得し、1 を足した数値をグループ情報 6 0 0 の更新番号に設定する。更新情報を格納する論理アドレスは、ポインタ情報 7 0 0 の更新情報最新アドレスであり、更新情報のサイズを足した数値をポインタ情報 7 0 0 の更新情報最新アドレスに設定する。ライトデータを格納する論理アドレスは、ポインタ情報 7 0 0 のライトデータ最新アドレスであり、ライトデータ最新アドレスにデータ量 A を足した数値をポインタ情報 7 0 0 のライトデータ最新アドレスに設定する。

【0 0 7 6】

ホストアダプタ A は、上記取得した数値とグループ番号、ライト命令を受信した時刻、ライト命令内の論理アドレス A、データ量 A を更新情報に設定する (ステップ 1 3 5 0、図 1 1 の 1 1 2 0)。例えば、図 6 に示すグループ情報 6 0 0、図 7 に示すポインタ情報 7 0 0 の状態で、グループ 1 に属する正論理ボリューム

ム 1 の記憶領域の先頭から 800 の位置にデータサイズ 100 のライト命令を受信した場合、図 22 に示す更新情報を作成する。グループ情報の更新番号は 5、ポインタ情報の更新情報最新アドレスは 600（更新情報のサイズは 100 とする）、ライトデータ最新アドレスは 2300 となる。

【0077】

(6) ホストアダプタ A は、ディスクアダプタ 120 に、ジャーナルの更新情報とライトデータを記憶装置 150 に書き込むことを命令し、正常終了する（ステップ 1360、図 11 の 1130、1140、1150）。

【0078】

(7) ジャーナル作成処理が、初期コピー処理内の処理の場合は、ステップ 1370 からの処理を行う。ホストアダプタ A は、ジャーナルが作成可能であることを調べる。ポインタ情報 700 を用い、更新情報領域の未使用領域の有無を調べる（ステップ 1370）。ポインタ情報 700 の更新情報最新アドレスと更新情報最古アドレスが等しい場合は、更新情報領域に未使用領域が存在しないため、ジャーナル作成失敗として処理を終了する（ステップ 1390）。本実施例で示した初期コピー処理の場合、ジャーナルのライトデータは、正論理ボリュームからリードし、ライトデータ領域は使用しないため、ライトデータ領域の未使用領域の確認は不要である。

【0079】

(8) ステップ 1370 の調べで、ジャーナルが作成可能である場合、ホストアダプタ A は、更新情報に設定する数値を取得し、更新情報をキャッシュメモリ 130 内に作成する。更新番号は、対象グループのグループ情報 600 から取得し、1 を足した数値をグループ情報 600 の更新番号に設定する。更新情報を格納する論理アドレスは、ポインタ情報 700 の更新情報最新アドレスの位置であり、更新情報のサイズを足した数値をポインタ情報 700 の更新情報最新アドレスに設定する。

【0080】

ホストアダプタ A は、上記取得した数値とグループ番号、本処理の開始時刻、初期コピー処理対象の論理アドレス、初期コピーの 1 回の処理量、ライトデータ

を格納したジャーナル論理ボリュームの論理アドレスに初期コピー処理対象の論理アドレスを設定する（ステップ1380、図11の1120）。

【0081】

(9) ホストアダプタAは、ディスクアダプタ120に、更新情報を記憶装置150に書き込むことを命令し、正常終了する（ステップ1385、図11の1140、1160）。

【0082】

上記説明では、更新情報をキャッシュメモリ130内に存在するように記載しているが、共有メモリ140内等に格納してもよい。

【0083】

ライトデータの記憶装置150への書き込みは、非同期、つまり、ステップ1360およびステップ1385の直後でなくともよい。ただし、ホストコンピュータ180が、論理アドレスAにライト命令を再び行った場合、ジャーナルのライトデータが上書きされるため、ホストコンピュータ180からライトデータを受信する前に、ジャーナルのライトデータは、更新情報のジャーナル論理ボリュームの論理アドレスに対応する記憶装置150に書き込む必要がある。もしくは、別のキャッシュメモリに退避し、後に更新情報のジャーナル論理ボリュームの論理アドレスに対応する記憶装置150に書き込みを行ってもよい。

【0084】

前述したジャーナル作成処理では、ジャーナルを記憶装置150に保存するとしていたが、ジャーナル用に予め一定量のキャッシュメモリ130を用意しておき、当該キャッシュメモリを全て使用してから、記憶装置150にジャーナルを保存してもよい。ジャーナル用のキャッシュメモリ量は、例えば、保守端末から指定する。

【0085】

リードライト処理220は、ディスクアダプタ120が、ホストアダプタ110もしくはディスクアダプタ120から命令を受け、実施する処理である。実施する処理は、指定されたキャッシュメモリ130のデータを指定された論理アドレスに対応する記憶装置150内の記憶領域に書き込む処理、指定された論理ア

ドレスに対応する記憶装置 150 内の記憶領域から指定されたキャッシュメモリ 130 にデータを読み込む処理等である。

【0086】

図 14 はジャーナルリード命令を受信した正記憶システム 100A のホストアダプタ A の動作（ジャーナルリード受信処理）を説明する図、図 15 はフローチャートである。以下、これらを用いて、正記憶システム 100A が、副記憶システム 100B からジャーナルリード命令を受信した場合の動作について説明する。

【0087】

(1) 正記憶システム 100A 内のホストアダプタ A は、副記憶システム 100B からアクセス命令を受信する。アクセス命令は、ジャーナルリード命令であることを示す識別子、命令対象のグループ番号、リトライ指示の有無を含んでいる。以下、アクセス命令内のグループ番号をグループ番号 A とする（ステップ 1220、図 14 の 1410）。

【0088】

(2) ホストアダプタ A は、グループ番号 A のグループ状態が“正常”であることを調べる（ステップ 1510）。ステップ 1510 の調べで、グループ状態が“正常”以外、例えば、“障害”の場合は、副記憶システム 100B にグループ状態を通知し、処理を終了する。副記憶システム 100B は、受信したグループ状態に応じて処理を行う。例えば、グループ状態が“障害”の場合は、ジャーナルリード処理を終了する（ステップ 1515）。

【0089】

(3) ステップ 1510 の調べで、グループ番号 A のグループ状態が“正常”の場合、ホストアダプタ A は、ジャーナル論理ボリュームの状態を調べる（ステップ 1520）。ステップ 1520 の調べで、ジャーナル論理ボリュームのボリューム状態が“正常”でない場合は、例えば、“障害”の場合は、グループ状態を“障害”に変更し、副記憶システム 100B にグループ状態を通知し、処理を終了する。副記憶システム 100B は、受信したグループ状態に応じて処理を行う。例えば、グループ状態が“障害”の場合は、ジャーナルリード処理を終了す

る（ステップ1525）。

【0090】

（4）ステップ1520の調べで、ジャーナル論理ボリュームのボリューム状態が“正常”の場合は、ジャーナルリード命令がリトライ指示かを調べる（ステップ1530）。

【0091】

（5）ステップ1530の調べで、ジャーナルリード命令がリトライ指示の場合、ホストアダプタAは、前回送信したジャーナルを再度、副記憶システム100Bに送信する。ホストアダプタAは、キャッシュメモリ130を確保し、ディスクアダプタに、ポインタ情報700のリトライ開始アドレスから、更新情報のサイズの情報をキャッシュメモリに読み込むことを命令する（図14の1420）。

【0092】

ディスクアダプタのリードライト処理は、記憶装置150から更新情報を読み込み、キャッシュメモリ130に保存し、ホストアダプタAに通知する（図14の1430）。

【0093】

ホストアダプタAは、更新情報のリード終了の通知を受け、更新情報から、ライトデータの論理アドレスおよびライトデータのサイズを取得し、キャッシュメモリ130を確保し、ディスクアダプタにライトデータをキャッシュメモリに読み込むことを命令する（ステップ1540、図14の1440）。

【0094】

ディスクアダプタのリードライト処理は、記憶装置150からライトデータを読み込み、キャッシュメモリ130に保存し、ホストアダプタAに通知する（図14の1450）。

【0095】

ホストアダプタAは、ライトデータのリード終了の通知を受け、更新情報とライトデータを副記憶システム100Bに送信し、ジャーナルを保持しているキャッシュメモリ130を開放し、処理を終了する（ステップ1545、図14の1

4 6 0)。

【0 0 9 6】

(6) ステップ1 5 3 0の調べで、リトライ指示でない場合、ホストアダプタ Aは、送信していないジャーナルが存在するかを調べ、存在すれば、ジャーナルを副記憶システム1 0 0 Bに送信する。ホストアダプタ Aは、ポインタ情報7 0 0のリード開始アドレスと更新情報最新アドレスを比較する(ステップ1 5 5 0)。

【0 0 9 7】

リード開始アドレスが更新情報最新アドレスと等しい場合は、全てのジャーナルを副記憶システム1 0 0 Bに送信済みであるため、副記憶システム1 0 0 Bに“ジャーナル無”を送信し(ステップ1 5 6 0)、前回のジャーナルリード命令の時に、副記憶システム1 0 0 Bに送信したジャーナルの記憶領域を開放する(ステップ1 5 9 0)。

【0 0 9 8】

ジャーナルの記憶領域の開放処理は、ポインタ情報7 0 0の更新情報最古アドレスに、リトライ開始アドレスを設定する。更新情報最古アドレスがライトデータ領域先頭アドレスとなった場合は、更新情報最古アドレスは0とする。ポインタ情報7 0 0のライトデータ最古アドレスは、前回のリードジャーナル命令に応じて送信したライトデータのサイズを足した数値に変更する。ライトデータ最古アドレスが、ジャーナル論理ボリュームの容量以上の論理アドレスとなった場合は、ライトデータ領域先頭アドレスを減じ、補正する。

【0 0 9 9】

(7) ステップ1 5 5 0の調べで、未送信のジャーナルが存在する場合、ホストアダプタ Aは、キャッシュメモリ1 3 0を確保し、ディスクアダプタにポインタ情報7 0 0のリード開始アドレスから、更新情報のサイズの情報をキャッシュメモリに読み込むことを命令する(図1 4の1 4 2 0)。

【0 1 0 0】

ディスクアダプタ Aのリードライト処理は、記憶装置1 5 0から更新情報を読み込み、キャッシュメモリ1 3 0に保存し、ホストアダプタに通知する(図1 4

の 1 4 3 0)。

【0 1 0 1】

ホストアダプタ A は、更新情報のリード終了の通知を受け、更新情報から、ライトデータの論理アドレスおよびライトデータのサイズを取得し、キャッシュメモリ 1 3 0 を確保し、ディスクアダプタ A にライトデータをキャッシュメモリに読み込むことを命令する (ステップ 1 5 7 0、図 1 4 の 1 4 4 0)。

【0 1 0 2】

ディスクアダプタ A のリードライト処理は、記憶装置 1 5 0 からライトデータを読み込み、キャッシュメモリ 1 3 0 に保存し、ホストアダプタに通知する (図 1 4 の 1 4 5 0)。

【0 1 0 3】

ホストアダプタ A は、ライトデータのリード終了の通知を受け、更新情報とライトデータを副記憶システム 1 0 0 B に送信 (ステップ 1 5 8 0) し、ジャーナルを保持しているキャッシュメモリ 1 3 0 を開放する (図 1 4 の 1 4 6 0)。そして、ポインタ情報 7 0 0 のリトライ開始アドレスにリード開始アドレスを設定し、リード開始アドレスに送信したジャーナルの更新情報サイズを足した数値を設定する。

【0 1 0 4】

(8) ホストアダプタ A は、前回のジャーナルリード命令の処理時に、副記憶システム 1 0 0 B に送信したジャーナルの記憶領域を開放する (ステップ 1 5 9 0)。

【0 1 0 5】

前述したジャーナルリード受信処理では、正記憶システム 1 0 0 A は、ジャーナルを一つずつ副記憶システム 1 0 0 B に送信していたが、複数同時に副記憶システム 1 0 0 B に送信してもよい。1 つのジャーナルリード命令で、送信するジャーナル数は、副記憶システム 1 0 0 B がジャーナルリード命令内に指定してもよいし、グループ登録の際等に、ユーザが正記憶システム 1 0 0 A もしくは、副記憶システム 1 0 0 B に指定してもよい。さらに、正記憶システム 1 0 0 A と副記憶システム 1 0 0 B の接続パス 2 0 0 の転送能力もしくは、負荷等により、動

的に1つのジャーナルリード命令で送信するジャーナル数を変更してもよい。また、ジャーナル数でなく、ジャーナルのライトデータのサイズを考慮し、ジャーナルの転送量を指定してもよい。

【0106】

前述したジャーナルリード受信処理では、ジャーナルを記憶装置150からキャッシュメモリ130に読み込んでいたが、キャッシュメモリ130に存在する場合は、当該処理は不要である。

【0107】

前述したジャーナルリード受信処理内のジャーナルの記憶領域の開放処理は、次のジャーナルリード命令の処理時としたが、副記憶システム100Bにジャーナルを送信した直後に開放してもよい。また、副記憶システム100Bが、ジャーナルリード命令内に開放してよい更新番号を設定し、正記憶システム100Aは、その指示に従って、ジャーナルの記憶領域を開放してもよい。

【0108】

図16はジャーナルリード命令処理240を説明する図、図17はフローチャート、図18はジャーナル格納処理のフローチャートである。以下、これらを用いて、副記憶システム100BのホストアダプタBが、正記憶システム100Aからジャーナルを読み出し、ジャーナル論理ボリュームに格納する動作について説明する。

【0109】

(1) 副記憶システム100B内のホストアダプタBは、ジャーナルを格納するキャッシュメモリ130を確保し、ジャーナルリード命令であることを示す識別子、命令対象の正記憶システム100Aのグループ番号、リトライ指示の有無を含むアクセス命令を正記憶システム100Aに送信する。以下、アクセス命令内のグループ番号をグループ番号Aとする（ステップ1700、図16の1610）。

【0110】

(2) ホストアダプタBは、正記憶システム100Aの応答およびジャーナルを受信する（図16の1620）。ホストアダプタBは応答を調べ、正記憶シス

テム 1 0 0 A からの応答が、“ジャーナル無”の場合は、正記憶システム 1 0 0 A には、指定したグループのジャーナルが存在しないため、一定時間後、正記憶システム 1 0 0 A にリードジャーナル命令を送信する（ステップ 1 7 2 0、1 7 2 5）。

【0 1 1 1】

（４）正記憶システム 1 0 0 A の応答が、“グループ状態は障害”もしくは“グループ状態は未使用”の場合は、副記憶システム 1 0 0 B のグループ状態を受信した状態に変更し、ジャーナルリード処理を終了する（ステップ 1 7 3 0、1 7 3 5）。

【0 1 1 2】

（５）正記憶システム 1 0 0 A の応答が、上記以外、つまり、正常終了の場合は、ジャーナル論理ボリュームのボリューム状態を調べる（ステップ 1 7 4 0）。ジャーナル論理ボリュームのボリューム状態が“異常”の場合は、ジャーナル論理ボリュームにジャーナルの格納が不可能なため、グループ状態を“異常”に変更し、処理を終了する（ステップ 1 7 4 5）。この場合、ジャーナル論理ボリュームを正常な論理ボリュームに変更する等を行い、グループの状態を正常に戻す。

【0 1 1 3】

（６）ステップ 1 7 4 0 の調べで、ジャーナル論理ボリュームのボリューム状態が“正常”の場合は、後述するジャーナル格納処理 1 8 0 0 を行う。ジャーナル格納処理 1 8 0 0 が正常に終了した場合は、次のジャーナルリード命令を送信する。もしくは、一定時間経過後、次のジャーナルリード命令を送信する（ステップ 1 7 6 0）。次のジャーナル命令を送信するタイミングは、一定の時間間隔で定期的に送信するものでもよく、受信したジャーナルの個数もしくは、接続線 2 0 0 の通信量、副記憶システム 1 0 0 B が保持しているジャーナルの記憶容量、副記憶システム 1 0 0 B の負荷等によって決めてもよい。さらに、正記憶システム 1 0 0 A が保持しているジャーナルの記憶容量、もしくは正記憶システム 1 0 0 A のポインタ情報を副記憶システム 1 0 0 B から読み出し、その数値に基づいて決めてもよい。上記情報の転送は、専用のコマンドで行ってもよいし、ジャ

ジャーナルリード命令の応答に含んでもよい。その後の処理は、ステップ 1 7 1 0 以降と同じである。

【0 1 1 4】

(7) ステップ 1 8 0 0 のジャーナル格納処理が正常に終了しない場合は、ジャーナル論理ボリュームの未使用領域が足りないため、受信したジャーナルを破棄し、一定時間後にリトライ指示のジャーナルリード命令を送信する（ステップ 1 7 5 5）。もしくは、ジャーナルをキャッシュメモリに保持しておき、一定時間後に、再度ジャーナル格納処理を行う。これは、後述するリストア処理 2 5 0 が行われることにより、一定時間後には、ジャーナル論理ボリュームに未使用領域が増える可能性があるためである。この方式の場合は、ジャーナルリード命令にリトライ指示の有無は不要である。

【0 1 1 5】

次に、図 1 8 に示すジャーナル格納処理 1 8 0 0 について説明する。

【0 1 1 6】

(1) ホストアダプタ B は、ジャーナルがジャーナル論理ボリュームに格納可能であるかを調べる。ポインタ情報 7 0 0 を用い、更新情報領域に未使用領域の有無を調べる（ステップ 1 8 1 0）。ポインタ情報 7 0 0 の更新情報最新アドレスと更新情報最古アドレスが等しい場合は、更新情報領域に未使用領域が存在しないため、ジャーナル作成失敗として処理を終了する（ステップ 1 8 2 0）。

【0 1 1 7】

(2) ステップ 1 8 1 0 の調べで、更新情報領域に未使用領域が存在する場合は、ポインタ情報 7 0 0 を用い、ライトデータ領域にライトデータが格納できるかを調べる（ステップ 1 8 3 0）。ライトデータ最新アドレスと受信したジャーナルのライトデータのデータ量の和が、ライトデータ最古アドレスと等しいもしくは大きい場合は、ライトデータ領域にライトデータを格納できないため、ジャーナル作成失敗として処理を終了する（ステップ 1 8 2 0）。

【0 1 1 8】

(3) ジャーナルが格納可能である場合、ホストアダプタ B は、受信した更新情報のグループ番号とジャーナル論理ボリュームの論理アドレスを変更する。グ

ループ番号は、副記憶システム 100B のグループ番号に変更し、ジャーナル論理ボリュームの論理アドレスはポインタ情報 700 のライトデータ最新アドレスに変更する。ホストアダプタ B は、ポインタ情報 700 の更新情報最新アドレスを更新情報最新アドレスに更新情報のサイズを足した数値に変更する。ホストアダプタ B は、ポインタ情報 700 のライトデータ最新アドレスを、ライトデータ最新アドレスにライトデータのサイズを足した数値に変更する（ステップ 1840）。

【0119】

（4）ホストアダプタ B は、ディスクアダプタ 120 に、更新情報とライトデータを記憶装置 150 に書き込むことを命令し、ジャーナル作成成功として処理を終了する（ステップ 1850、図 16 の 1630）。その後、ディスクアダプタ 120 は、リードライト処理により、記憶装置 150 に更新情報とライトデータを書き込み、キャッシュメモリ 130 を開放する（図 16 の 1640）。

【0120】

前述したジャーナル格納処理では、ジャーナルを記憶装置 150 に保存するとしていたが、ジャーナル用に予め一定量のキャッシュメモリ 130 を用意しておき、当該キャッシュメモリを全て使用してから、記憶装置 150 にジャーナルを保存してもよい。ジャーナル用のキャッシュメモリ量は、例えば、保守端末から指定する。

【0121】

図 19 はリストア処理 250 を説明する図、図 20 はフローチャートである。以下、これらを用いて、副記憶システム 100B のホストアダプタ B が、ジャーナルを利用し、データの更新を行う動作について説明する。リストア処理 250 は副記憶システム 100B のディスクアダプタ 120 が処理を行ってもよい。

【0122】

（1）ホストアダプタ B は、グループ番号 B のグループ状態が“正常”であるかを調べる（ステップ 2010）。ステップ 2010 の調べで、グループ状態が“正常”以外、例えば、“障害”の場合は、リストア処理を終了する（ステップ 2015）。

【0123】

(2) ステップ2010の調べで、グループ状態が“正常”の場合は、ジャーナル論理ボリュームのボリューム状態を調べる(ステップ2020)。ステップ2020の調べで、ジャーナル論理ボリュームのボリューム状態が、“異常”の場合は、アクセス不可能なため、グループ状態を“異常”に変更し、処理を終了する(ステップ2025)。

【0124】

(3) ステップ2020の調べで、ジャーナル論理ボリュームのボリューム状態が、“正常”の場合は、リストア対象のジャーナルが存在するかを調べる。ホストアダプタBは、ポインタ情報700の更新情報最古アドレスと更新情報最新アドレスを取得する。更新情報最古アドレスと更新情報最新アドレスが等しい場合、ジャーナルは存在しないため、リストア処理は一旦終了し、一定時間後、リストア処理を再開する(ステップ2030)。

【0125】

(4) ステップ2030の調べで、リストア対象のジャーナルが存在する場合、最古(最小)の更新番号を持つジャーナルに対して次の処理を行う。最古(最小)の更新番号を持つジャーナルの更新情報は、ポインタ情報700の更新情報最古アドレスから保存されている。ホストアダプタBは、キャッシュメモリ130を確保し、ディスクアダプタに更新情報最古アドレスから、更新情報のサイズの情報をキャッシュメモリ130に読み込むことを命令する(図19の1910)。

【0126】

ディスクアダプタのリードライト処理は、記憶装置150から更新情報を読み込み、キャッシュメモリ130に保存し、ホストアダプタBに通知する(図19の1920)。

【0127】

ホストアダプタBは、更新情報のリード終了の通知を受け、更新情報から、ライトデータの論理アドレスおよびライトデータのサイズを取得し、キャッシュメモリ130を確保し、ディスクアダプタにライトデータをキャッシュメモリに読

み込むことを命令する（図 1 9 の 1 9 3 0）。

【0 1 2 8】

ディスクアダプタのリードライト処理は、記憶装置 1 5 0 からライトデータを読み込み、キャッシュメモリ 1 3 0 に保存し、ホストアダプタに通知する（ステップ 2 0 4 0、図 1 9 の 1 9 4 0）。

【0 1 2 9】

（5）ホストアダプタ B は、更新情報から更新する副論理ボリュームの論理アドレスを求め、ディスクアダプタに副論理ボリュームにライトデータを書き込むことを命令する（ステップ 2 0 5 0、図 1 9 の 1 9 5 0）。ディスクアダプタのリードライト処理は、副論理ボリュームの論理アドレスに対応する記憶装置 1 5 0 にデータを書き込み、キャッシュメモリ 1 3 0 を開放し、ホストアダプタに通知する（図 1 9 の 1 9 6 0）。

【0 1 3 0】

（6）ホストアダプタ B は、ディスクアダプタのライト処理完了の通知を受け、ジャーナルの記憶領域を開放する。ジャーナルの記憶領域の開放処理は、ポインタ情報 7 0 0 の更新情報最古アドレスを更新情報のサイズを足した数値に変更する。更新情報最古アドレスが、ライトデータ領域先頭アドレスとなった場合は、ライトデータ領域先頭アドレスは 0 とする。ポインタ情報 7 0 0 のライトデータ最古アドレスは、ライトデータのサイズを足した数値に変更する。ライトデータ最古アドレスが、ジャーナル論理ボリュームの容量以上の論理アドレスとなった場合は、ライトデータ領域先頭アドレスを減じ、補正する。その後、ホストアダプタ B は、次のリストア処理を開始する（ステップ 2 0 6 0）。

【0 1 3 1】

前述したリストア処理 2 5 0 では、記憶装置 1 5 0 からキャッシュメモリ 1 3 0 にジャーナルを読み込んでいたが、キャッシュメモリ 1 3 0 に存在する場合は、当該処理は不要である。

【0 1 3 2】

前述したジャーナルリード受信処理とジャーナルリード命令処理 2 4 0 では、正記憶システム 1 0 0 A が送信するジャーナルをポインタ情報 7 0 0 により決め

ていたが、副記憶システム 100B が送信するジャーナルを決めてもよい。例えば、ジャーナルリード命令に更新番号を追加する。この場合、ジャーナルリード受信処理にて、副記憶システム 100B が指定した更新番号の更新情報の論理アドレスを求めるために、正記憶システム 100A の共有メモリ 140 内に、更新番号から更新情報を格納した論理アドレスを求めるテーブルもしくは検索方法を設ける。

【0133】

前述したジャーナルリード受信処理とジャーナルリード命令処理 240 では、ジャーナルリード命令を用いていたが、通常のリード命令を用いてもよい。例えば、正記憶システム 100A のグループ情報 600 とポインタ情報 700 を予め副記憶システム 100B に転送しておき、副記憶システム 100B は、正記憶システム 100A のジャーナル論理ボリュームのデータ（つまり、ジャーナル）をリードする。

【0134】

前述したジャーナルリード受信処理では、更新番号の順に、正記憶システム 100A から副記憶システム 100B にジャーナルを送信すると説明したが、更新番号の順に送信せずともよい。また、正記憶システム 100A から副記憶システム 100B に複数のジャーナルリード命令を送信してもよい。この場合、リストア処理にて更新番号順にジャーナルを処理するために、副記憶システム 100B に、更新番号から更新情報を格納した論理アドレスを求めるテーブルもしくは検索方法を設ける。

【0135】

前述した本発明のデータ処理システムでは、正記憶システムがジャーナルを取得し、副記憶システムがデータの複製を行う。これにより、正記憶システムに接続したホストコンピュータは、データの複製に関する負荷を負わない。さらに、正記憶システムと副記憶システム間でジャーナルを転送することにより、正記憶システムと正記憶システムに接続したホストコンピュータの通信線を使用しない。

【0136】

図 23 は、第二の実施形態の論理的な構成を示す図である。

【0137】

ホストコンピュータ 180 と記憶システム 100C を接続パス 190 により接続し、記憶システム 100C と正記憶システム 100A を接続パス 200 により接続し、正記憶システム 100A と副記憶システム 100B を接続パス 200 により接続した構成である。記憶システム 100C は、記憶システム 100C の論理ボリューム (ORG1) へのデータ更新時、論理ボリューム (ORG1) のデータ更新と正記憶システム 100A 内の論理ボリューム (DATA1) のデータ更新を行う。

【0138】

正記憶システム 100A は、第 1 の実施例で説明した通り、正論理ボリューム (DATA1) へのデータ更新時、前述した命令受信処理 210 およびリードライト処理 220 によって、ジャーナル論理ボリューム (JNL1) にジャーナルの保存を行う (2310)。

【0139】

副記憶システム 100B は、前述したジャーナルリード処理 240 によって、正記憶システム 100A からジャーナルをリードし、リードライト処理 220 によって、ジャーナル論理ボリューム (JNL2) にジャーナルを保存する (2320)。

【0140】

正記憶システム 100A は、副記憶システム 100B からジャーナルをリードする命令を受信すると、命令受信処理 210 およびリードライト処理 220 によって、ジャーナル論理ボリューム (JNL1) からジャーナルを読み出し、副記憶システム 100B に送信する (2320)。

【0141】

副記憶システム 100B は、前述したリストア処理 250 およびリードライト処理 220 によって、更新番号に従い、ジャーナル論理ボリューム (JNL2) からジャーナルを読み出し、正論理ボリューム (DATA1) の複製である副論理ボリューム (COPY1) のデータを更新する (2330)。このように、更新

番号の順にデータを更新することにより、論理ボリューム間のデータの整合性を保つことが可能となる。

【0142】

前述した本発明のデータ処理システムでは、正記憶システムがジャーナルを取得し、ジャーナル専用の記憶領域に格納する。さらに、副記憶システムは、正記憶システムから受信したジャーナルをジャーナル専用の記憶領域に格納する。ジャーナル専用の記憶領域はデータ複製対象の記憶領域より少なくすることが可能であり、より少ない記憶容量で、副記憶システムに正記憶システムのデータの複製が可能となる。

【0143】

図24は、第三の実施形態の論理的な構成を示す図である。

【0144】

ホストコンピュータ180と記憶システム100Cを接続パス190により接続し、記憶システム100Cと正記憶システム100Aを接続パス200により接続し、正記憶システム100Aと副記憶システム100Bを接続パス200により接続した構成である。記憶システム100Cは、従来技術で説明した通り、記憶システム100Cの論理ボリューム（ORG1）へのデータ更新時、論理ボリューム（ORG1）のデータ更新と正記憶システム100A内の論理ボリューム（DATA1）のデータ更新を行う。

【0145】

正記憶システム100Aは、記憶システム100Cに対し、正論理ボリューム（DATA1）があるように見せるが、実際の記憶領域、つまり記憶装置150は割り当てない。例えば、ボリューム情報400の物理アドレスに記憶装置150を割り当てていないことを示す数値を設定する。正記憶システム100Aは、記憶システム100Cから正論理ボリューム（DATA1）のデータへのライト命令受信時、前述した命令受信処理210内のステップ1270の処理を行わず、ジャーナル論理ボリューム（JNL1）にジャーナルの保存のみを行う（2410）。

【0146】

副記憶システム 1 0 0 B は、前述したジャーナルリード処理 2 4 0 によって、正記憶システム 1 0 0 A からジャーナルをリードし、リードライト処理 2 2 0 によって、ジャーナル論理ボリューム（JNL 2）にジャーナルを保存する（2 4 2 0）。

【0 1 4 7】

正記憶システム 1 0 0 A は、副記憶システム 1 0 0 B からジャーナルをリードする命令を受信すると、命令受信処理 2 1 0 およびリードライト処理 2 2 0 によって、ジャーナル論理ボリューム（JNL 1）からジャーナルを読み出し、記憶システム 1 0 0 B に送信する（2 4 2 0）。

【0 1 4 8】

副記憶システム 1 0 0 B は、前述したリストア処理 2 5 0 およびリードライト処理 2 2 0 によって、更新番号に従い、ジャーナル論理ボリューム（JNL 2）からジャーナルを読み出し、論理ボリューム（ORG 1）の複製である副論理ボリューム（COPY 1）のデータを更新する（2 4 3 0）。このように、更新番号の順にデータを更新することにより、論理ボリューム間のデータの整合性を保つことが可能となる。

【0 1 4 9】

前述した本発明のデータ処理システムでは、記憶システム 1 0 0 C もしくは、記憶システム 1 0 0 C に接続したホストコンピュータ 1 8 0 に障害が生じた場合、副記憶システム 1 0 0 B の論理ボリューム（COPY 1）に対し、正記憶システム 1 0 0 A 内のジャーナル（JNL 1）を反映することにより、記憶システム 1 0 0 B に接続したホストコンピュータにより最新データの参照、更新が可能となる。さらに、正記憶システム 1 0 0 A にデータの複製を保持せず、ジャーナルのみを格納することで、データ複製に必要な記憶容量が少なくすることが可能となる。

【0 1 5 0】

以上、本発明者によってなされた発明を実施例の形態に基づき具体的に説明したが、本発明は前記実施の形態に限定されるものでなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。

【0 1 5 1】

【発明の効果】

本発明によれば、記憶システムの上位の計算機に影響を与えず、複数の記憶システム間でデータ転送又はデータの複製をすることが可能な記憶システムを提供できる。さらに、記憶システムと計算機との間の通信にも影響を与えない記憶システムを提供できる。

【0 1 5 2】

さらに、本発明によれば、複数の記憶システム内に保持するデータ格納領域を少なくすることができる。さらに、複数の記憶システムの上位の計算機の業務に影響を与えることのないように、高速かつ効率的に複数の記憶システム間でデータ転送又はデータの複製をすることができる。

【図面の簡単な説明】

【図 1】 本発明の一実施形態の論理的な構成を示すブロック図である。

【図 2】 本発明の一実施形態の記憶システムのブロック図である。

【図 3】 本発明の一実施形態の更新情報とライトデータの関係を示す図である。

【図 4】 本発明の一実施形態のボリューム情報の例を示す図である。

【図 5】 本発明の一実施形態のペア情報の例を示す図である。

【図 6】 本発明の一実施形態のグループ情報の例を示す図である。

【図 7】 本発明の一実施形態のポインタ情報の例を示す図である。

【図 8】 本発明の一実施形態のジャーナル論理ボリュームの構造を示す図である。

【図 9】 本発明の一実施形態のデータの複製を開始する手順を示すフローチャートである。

【図 1 0】 本発明の一実施形態の初期コピー処理を示すフローチャートである。

【図 1 1】 本発明の一実施形態の命令受信処理を示す図である。

【図 1 2】 本発明の一実施形態の命令受信処理のフローチャートである。

【図 1 3】 本発明の一実施形態のジャーナル作成処理のフローチャートである。

【図 1 4】本発明の一実施形態のジャーナルリード受信処理を説明する図である。

【図 1 5】本発明の一実施形態のジャーナルリード受信処理のフローチャートである。

【図 1 6】本発明の一実施形態のジャーナルリード命令処理を説明する図である。

【図 1 7】本発明の一実施形態のジャーナルリード命令処理のフローチャートである。

【図 1 8】本発明の一実施形態のジャーナル格納処理のフローチャートである。

【図 1 9】本発明の一実施形態のリストア処理を説明する図である。

【図 2 0】本発明の一実施形態のリストア処理のフローチャートである。

【図 2 1】本発明の一実施形態の更新情報の例を説明する図である。

【図 2 2】本発明の一実施形態のジャーナル作成処理時の更新情報の例を説明する図である。

【図 2 3】本発明の第 2 の実施形態を示す図である。

【図 2 4】本発明の第 3 の実施形態を示す図である。

【符号の説明】

1 0 0 記憶システム

1 1 0 ホストアダプタ

1 2 0 ディスクアダプタ

1 3 0 キャッシュメモリ

1 4 0 管理メモリ

1 5 0 記憶装置

1 6 0 コモンパス

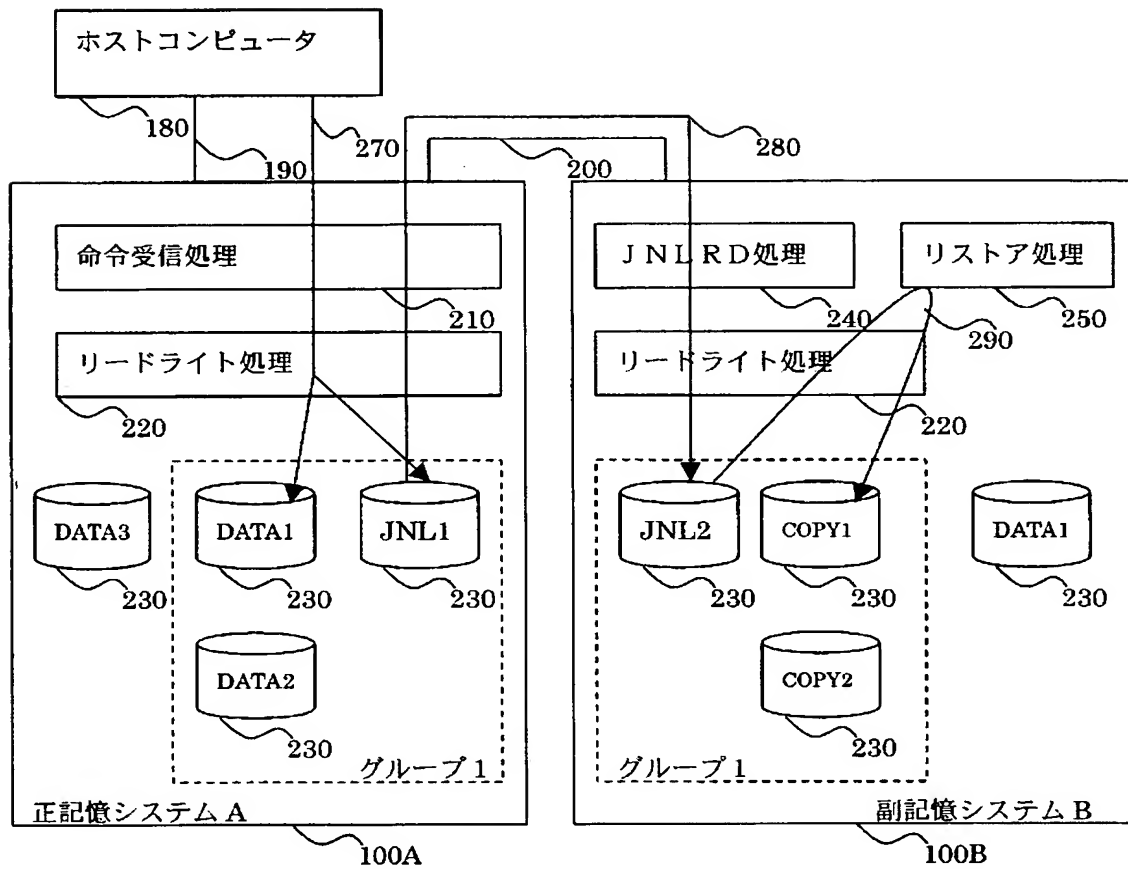
1 7 0 ディスクアダプタと記憶装置間の接続線

1 8 0 ホストコンピュータ

【書類名】 図面

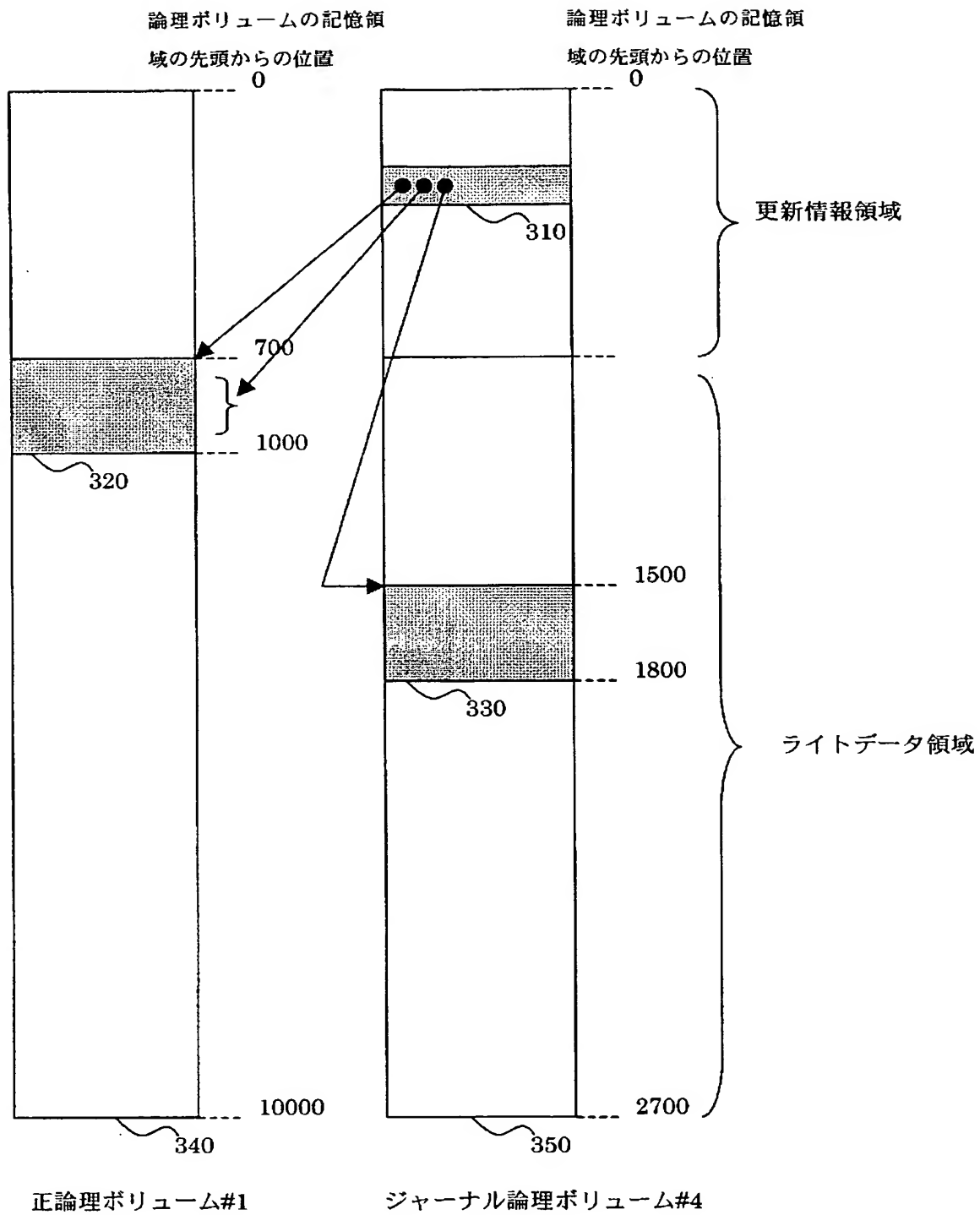
【図 1】

図 1



【図 3】

図 3



【図 4】

図 4

論理 ポリューム 番号	ポリューム 状態	フォー マット 形式	容量	ペア番号	物理アドレス	
					記憶装 置番号	先頭から位 置
1	正	OPEN3	3	1	1	0
2	正	OPEN6	6	2	1	3
3	未使用	OPEN6	6	0	1	9
4	正常	OPEN9	9	0	2	0
5	正常	OPEN3	3	0	2	9
6	未使用	OPEN6	6	0	2	12

400 ポリューム情報

【図 5】

図 5

ペア番 号	ペア状態	正記憶シス テム番号	正論理ポ リューム 番号	副記憶シス テム番号	副論理ポ リューム 番号	グルー プ番号	コピー 済みア ドレス
1	正常	1	1	2	1	1	0
2	正常	1	2	2	3	1	0
3	未使用	0	0	0	0	0	0
4	未使用	0	0	0	0	0	0
5	未使用	0	0	0	0	0	0

500 ペア情報

【図 6】

図 6

グルー プ 番号	グループ状態	ペア集合	ジャーナル論理 ポリューム番号	更新番号
1	正常	1,2	4	1
2	未使用	0	0	0

600 グループ情報

【図 7】

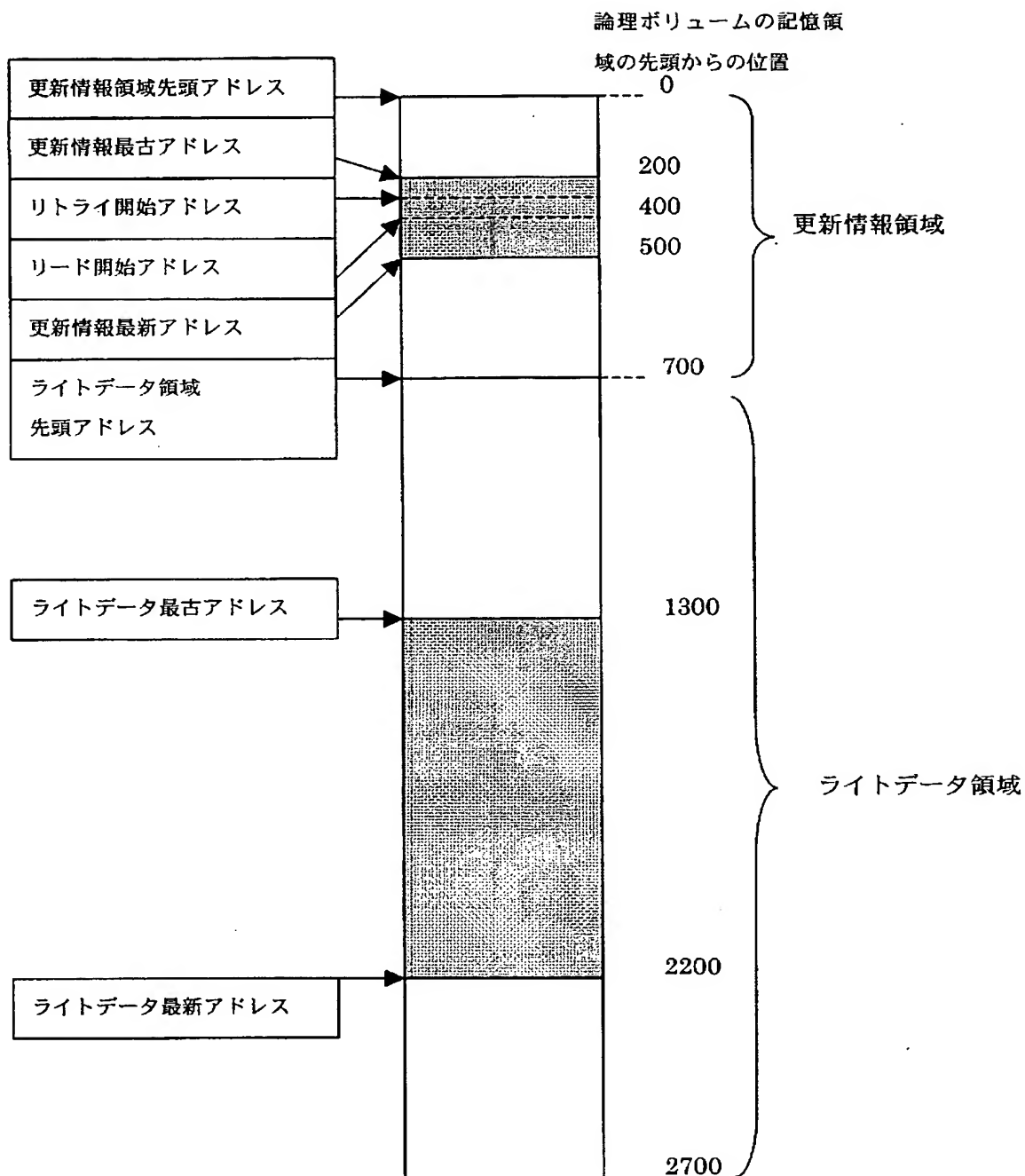
図 7

	論理アドレス	
	論理ボリューム番号	論理ボリュームの記憶領域の先頭からの位置
更新情報領域先頭アドレス	4	0
ライトデータ領域先頭アドレス	4	700
更新情報最新アドレス	4	500
更新情報最古アドレス	4	200
ライトデータ最新アドレス	4	2200
ライトデータ最古アドレス	4	1300
リード開始アドレス	4	400
リトライ開始アドレス	4	300

700 ポインタ情報

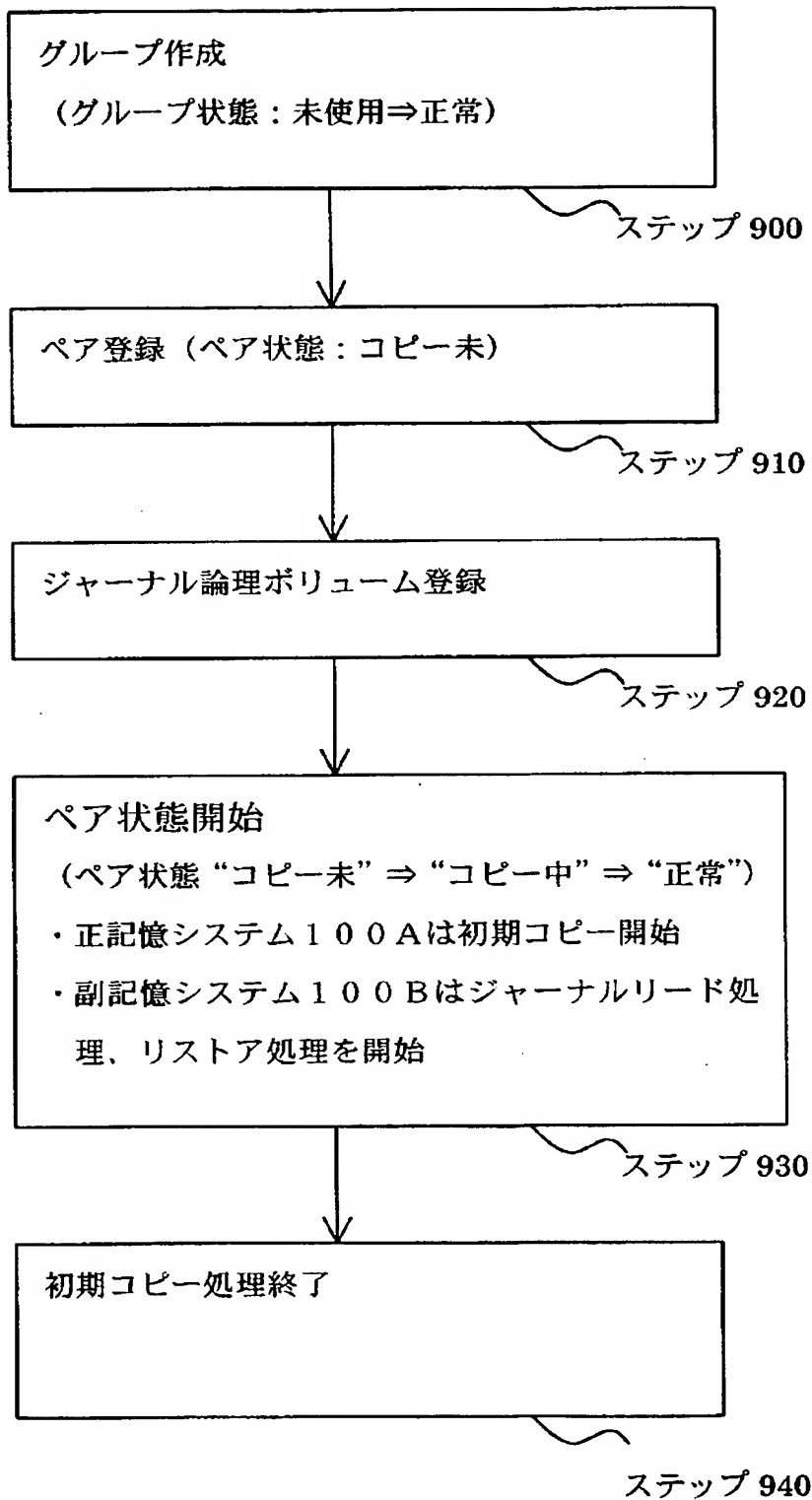
【図 8】

図 8



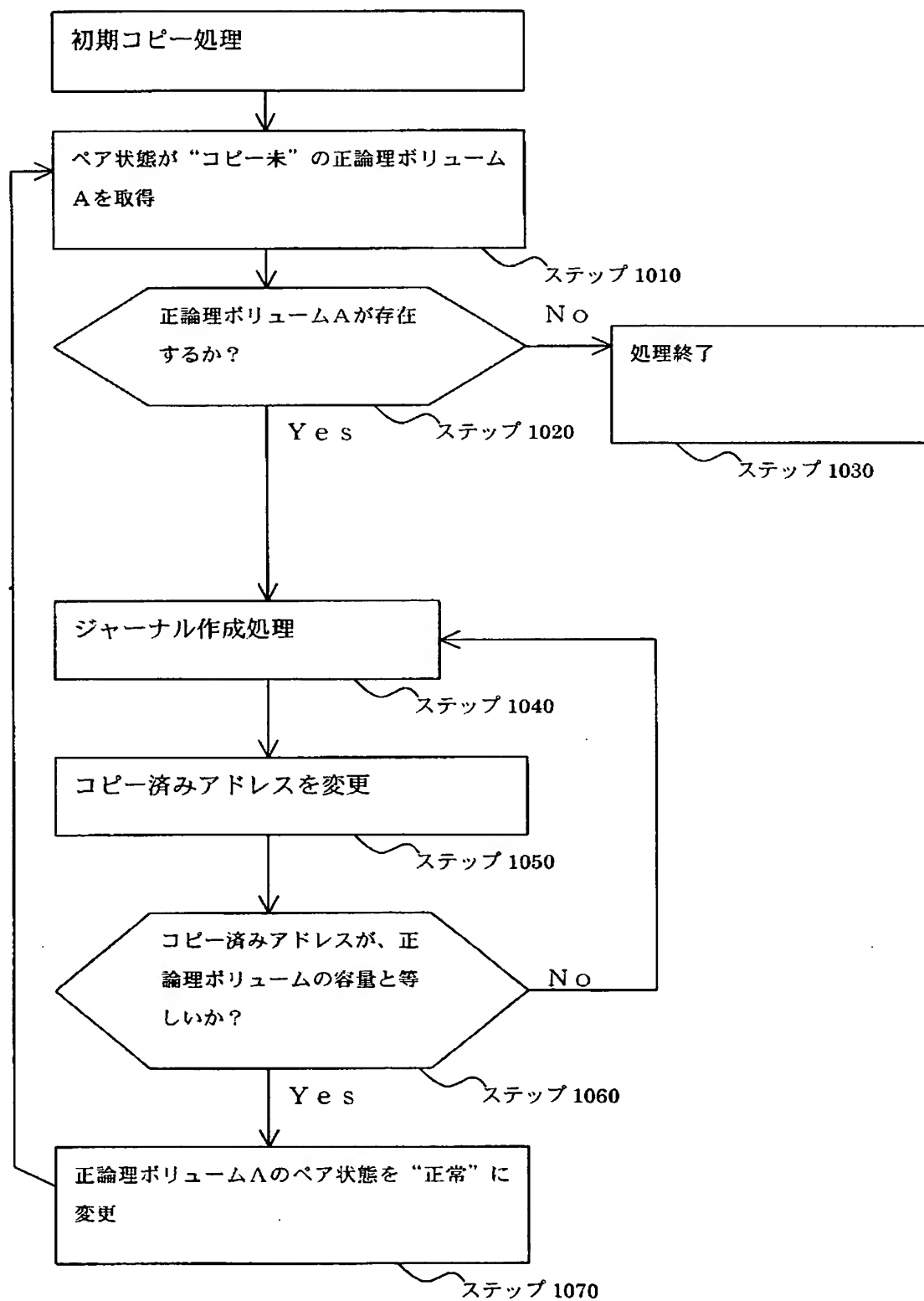
【図 9】

図 9



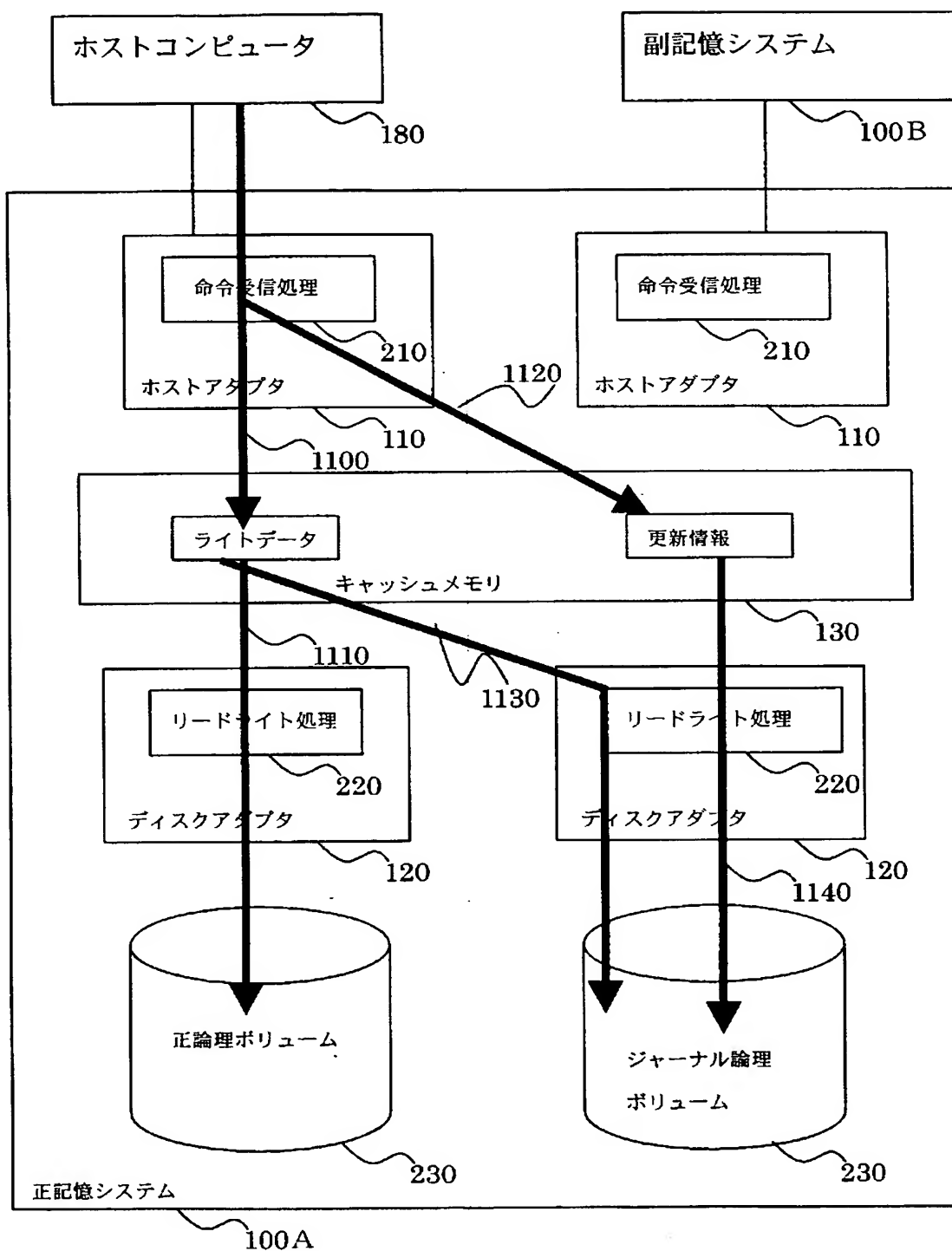
【図 10】

図 10

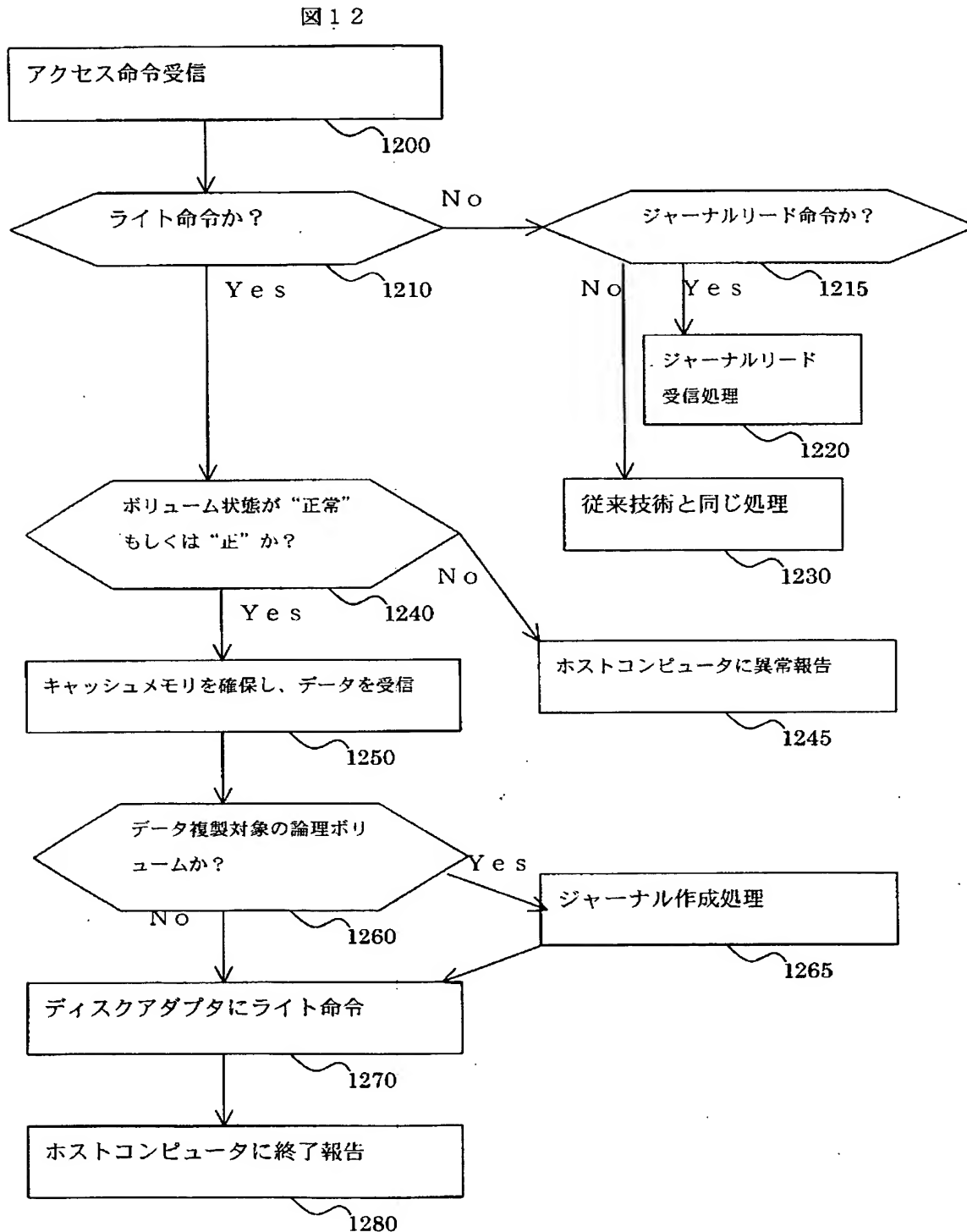


【図 11】

図 11

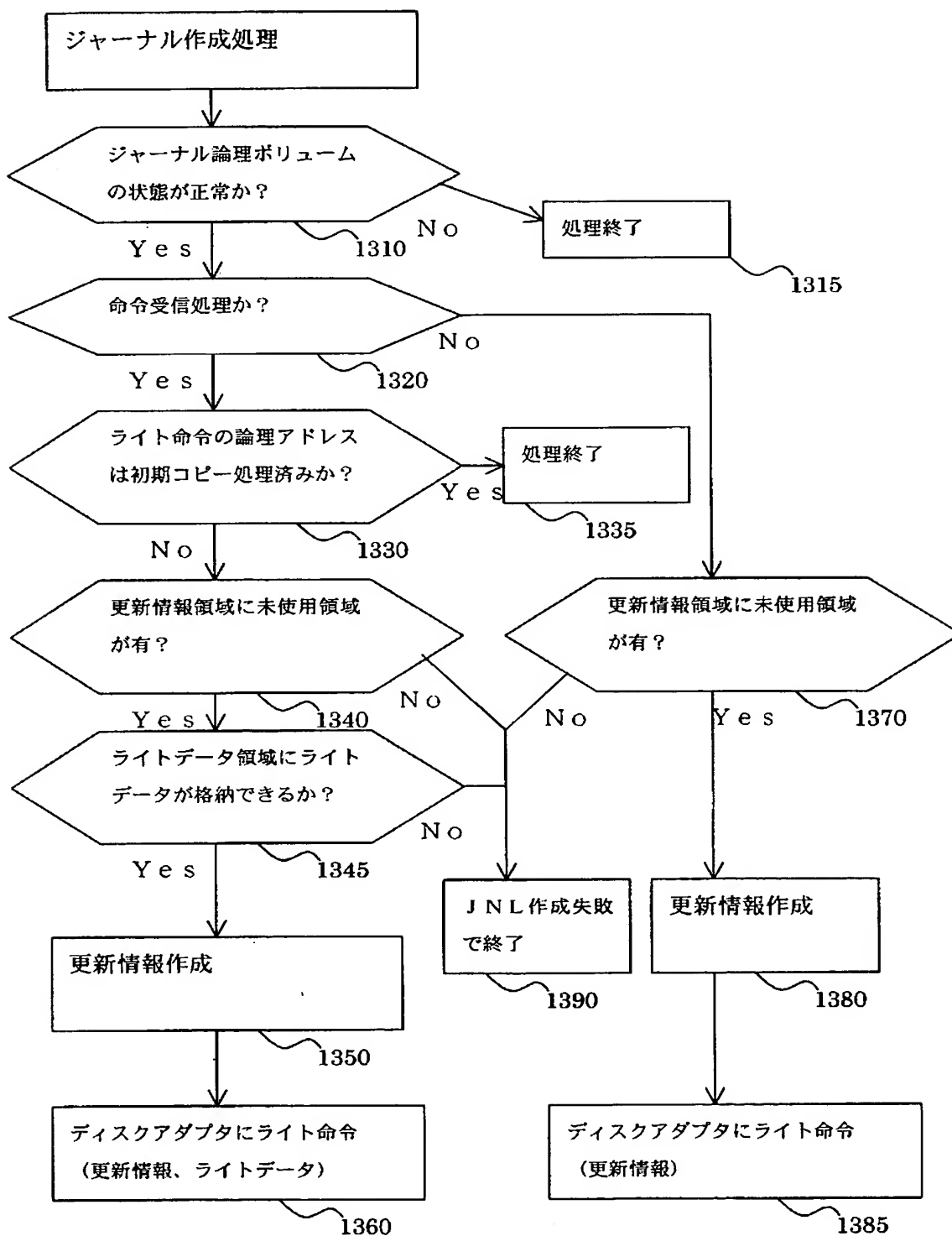


【図12】



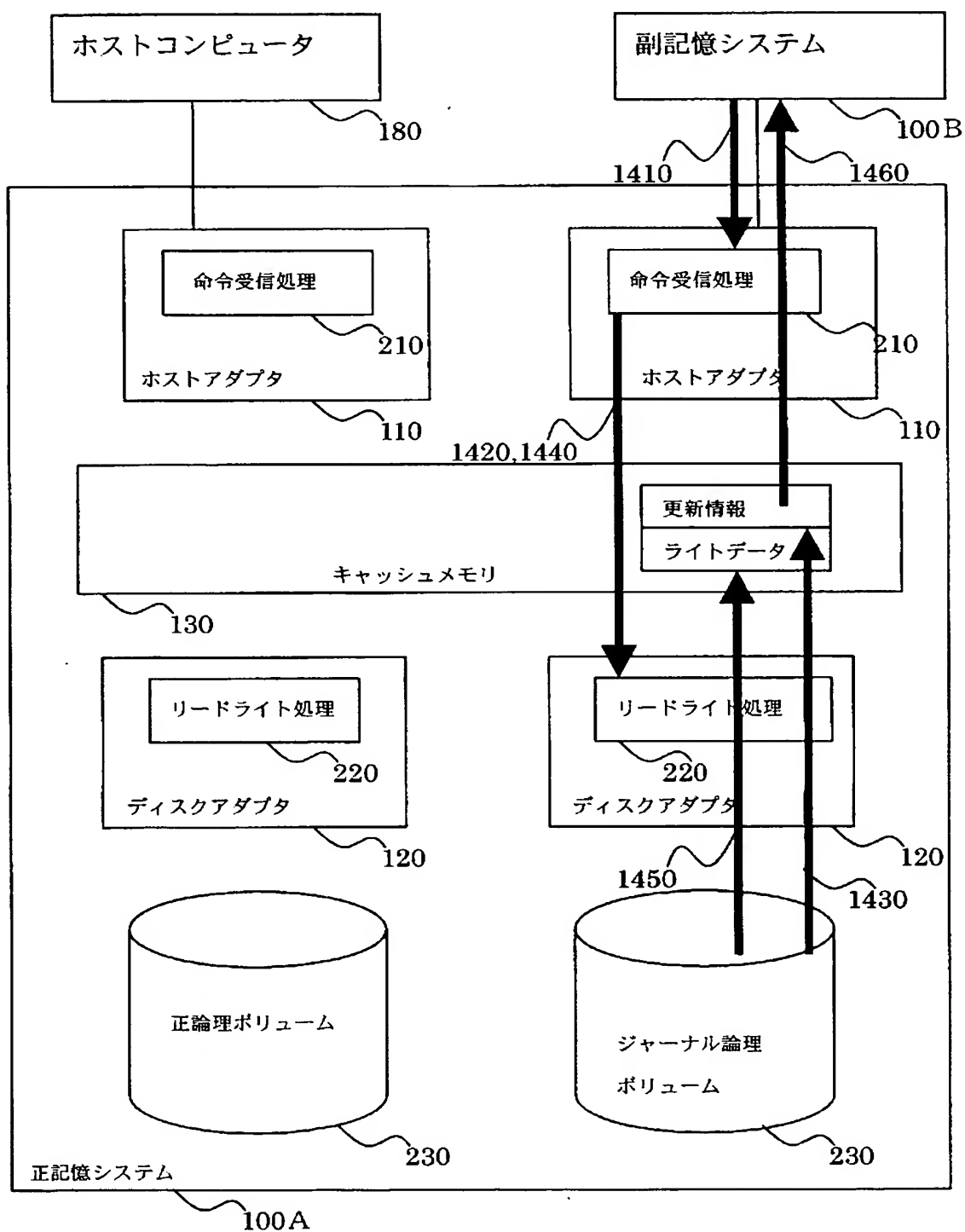
【図 13】

図 13

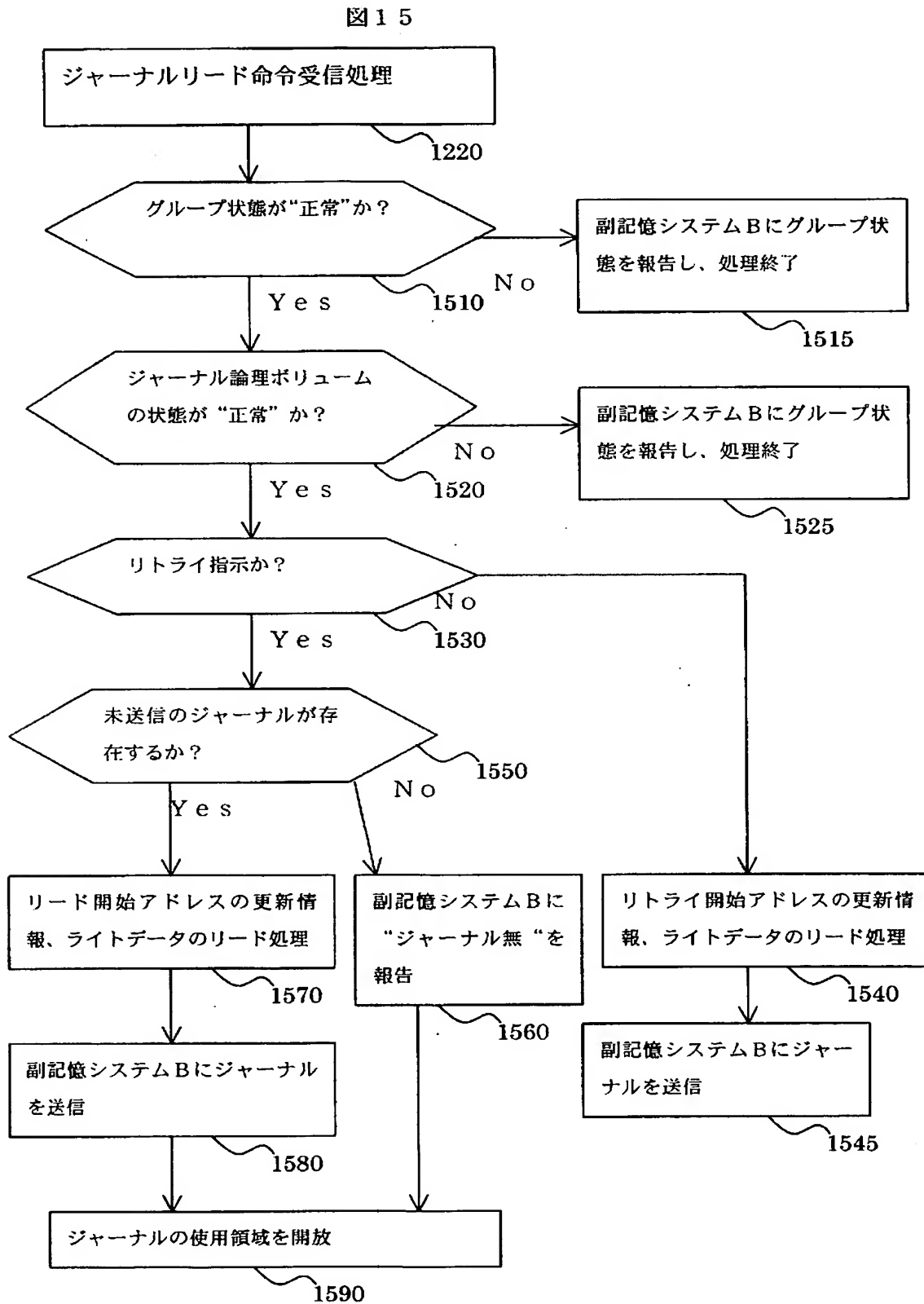


【図 14】

図 14

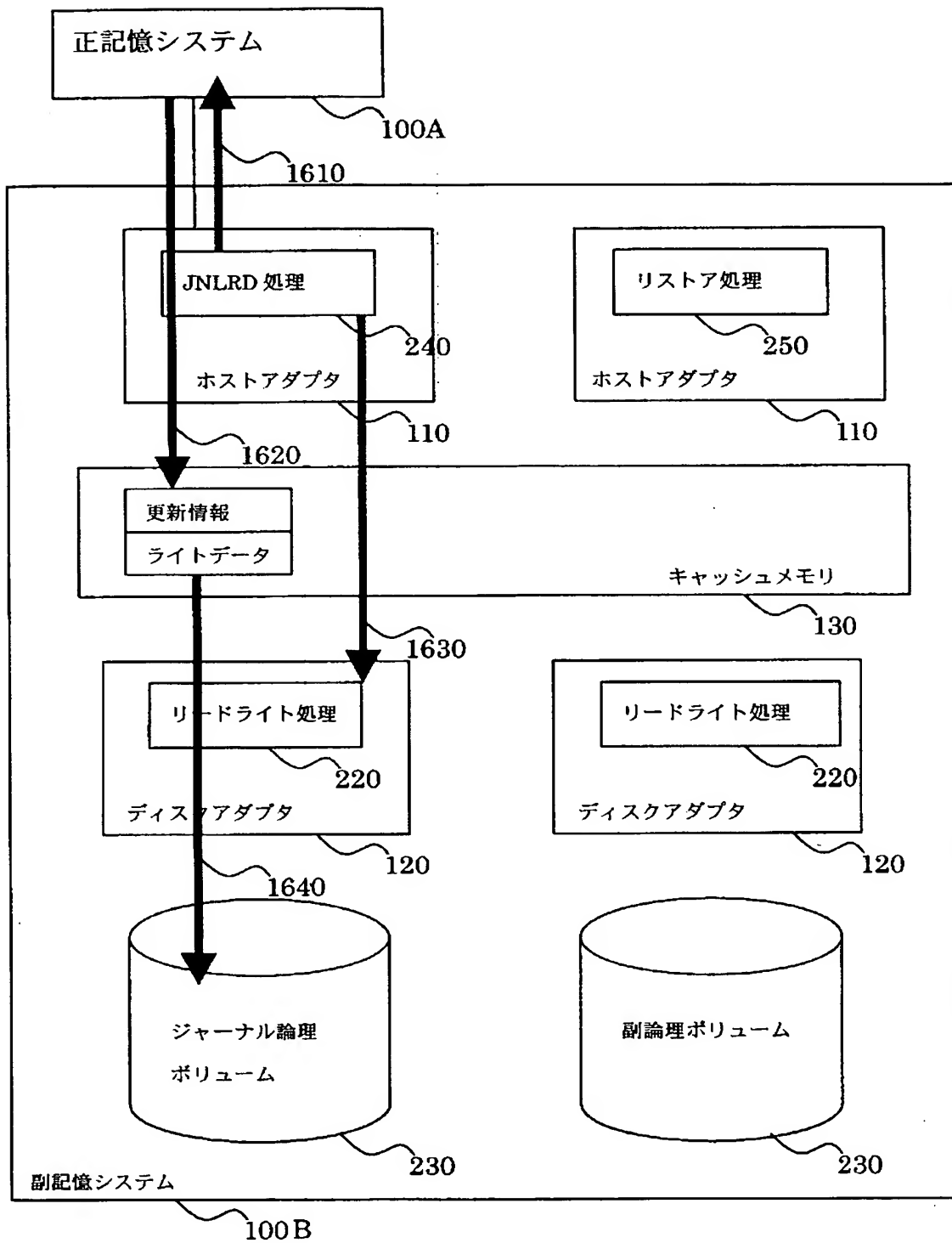


【図 15】



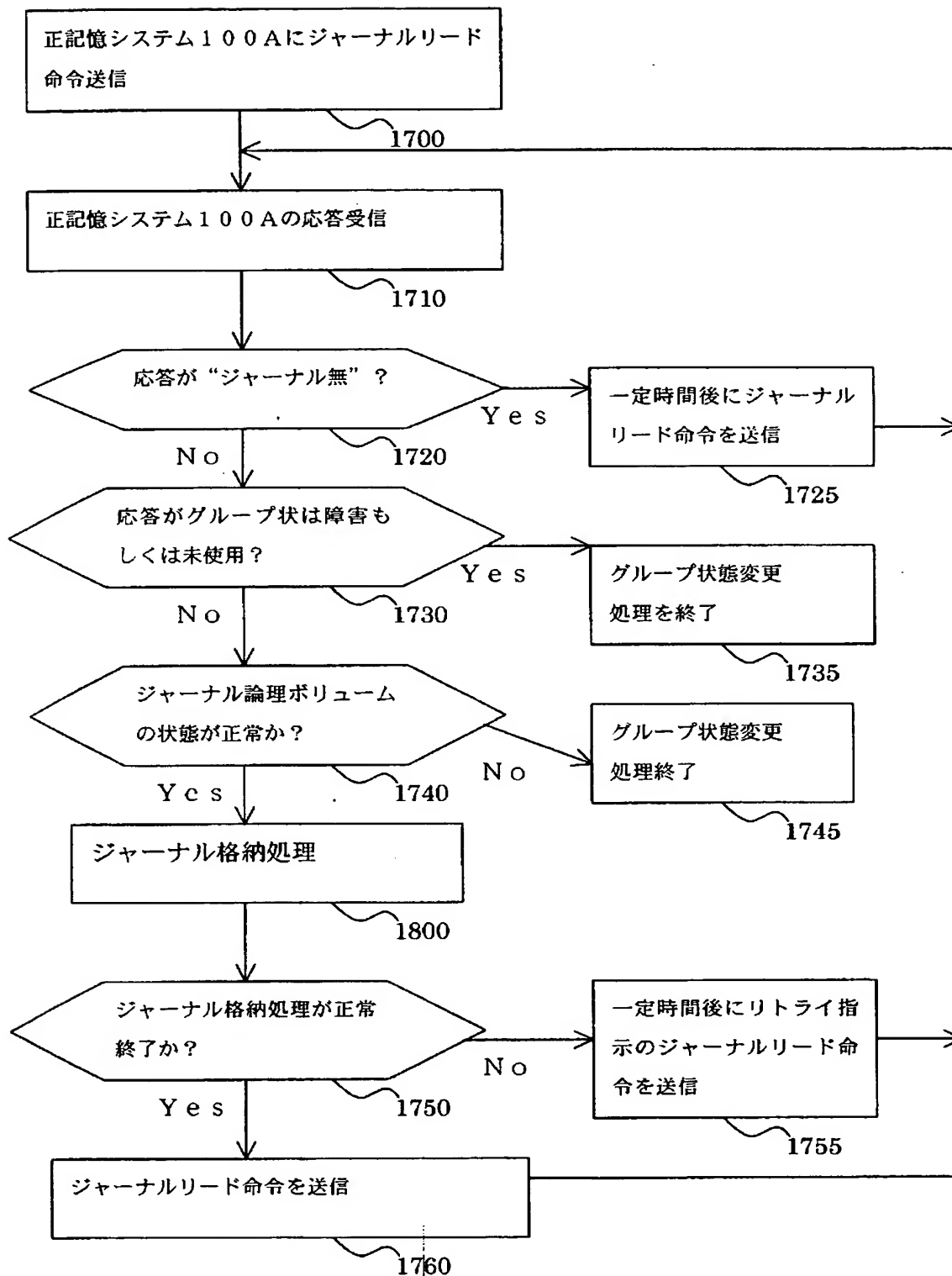
【図 16】

図 16



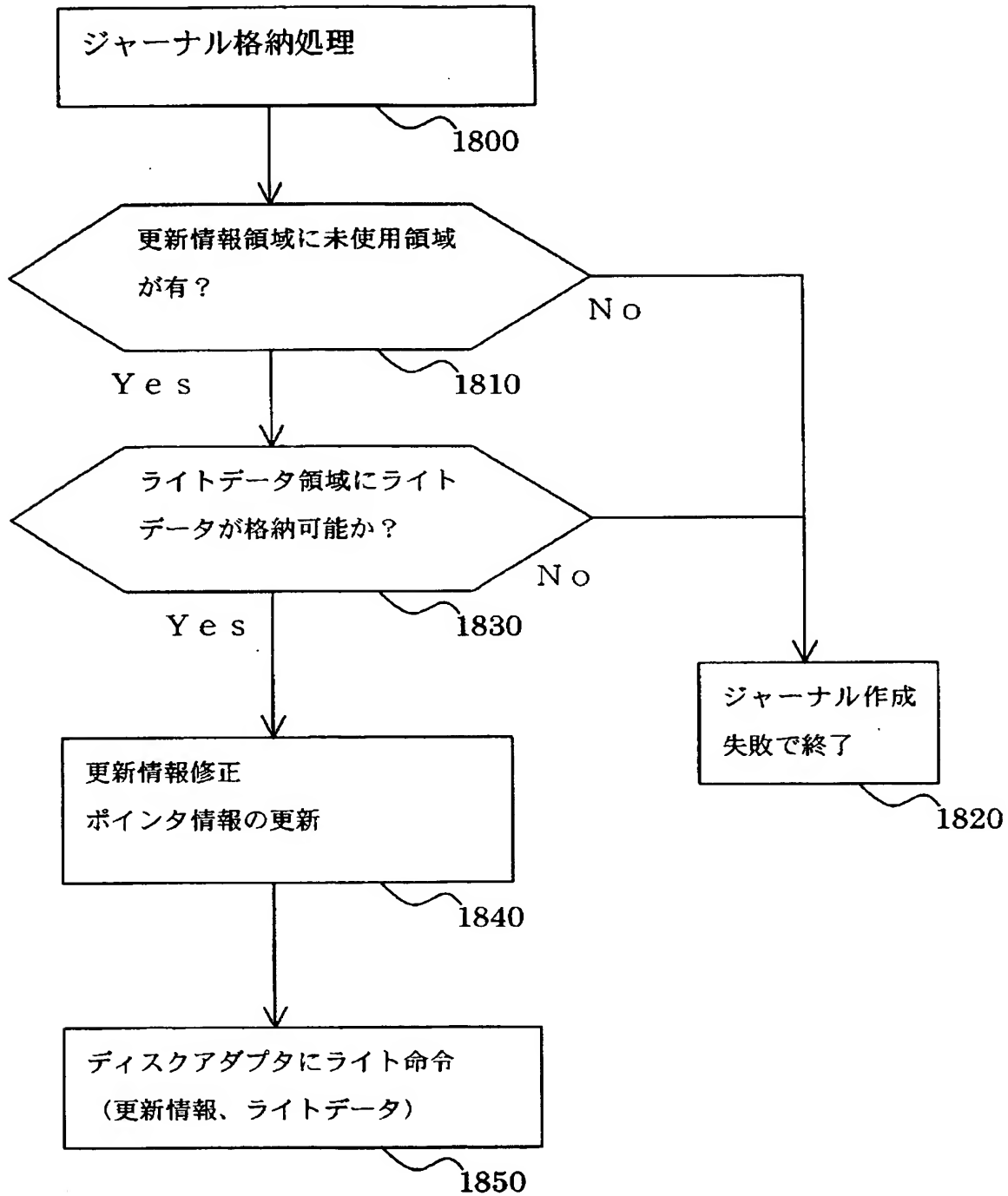
【図17】

図17



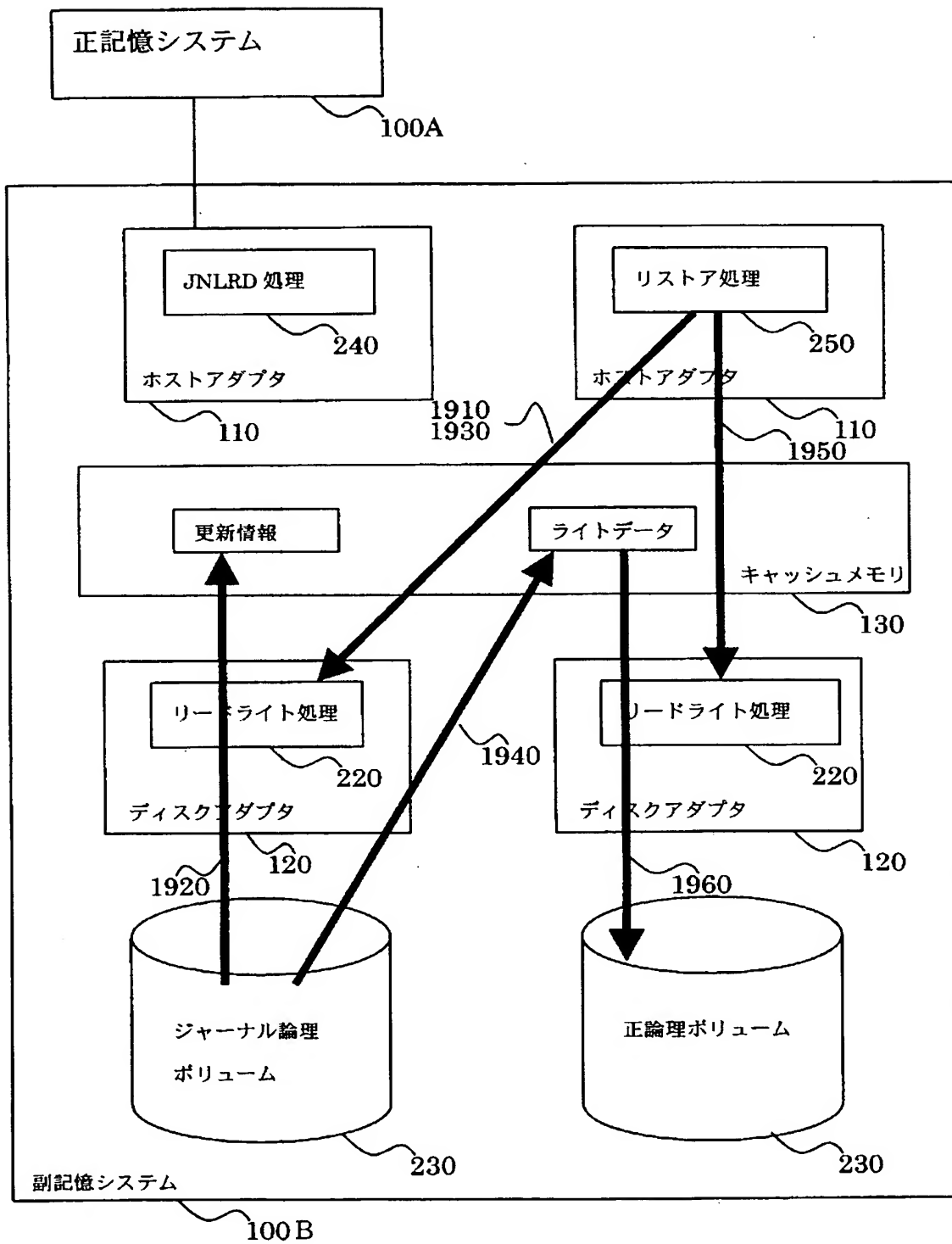
【図 18】

図 18



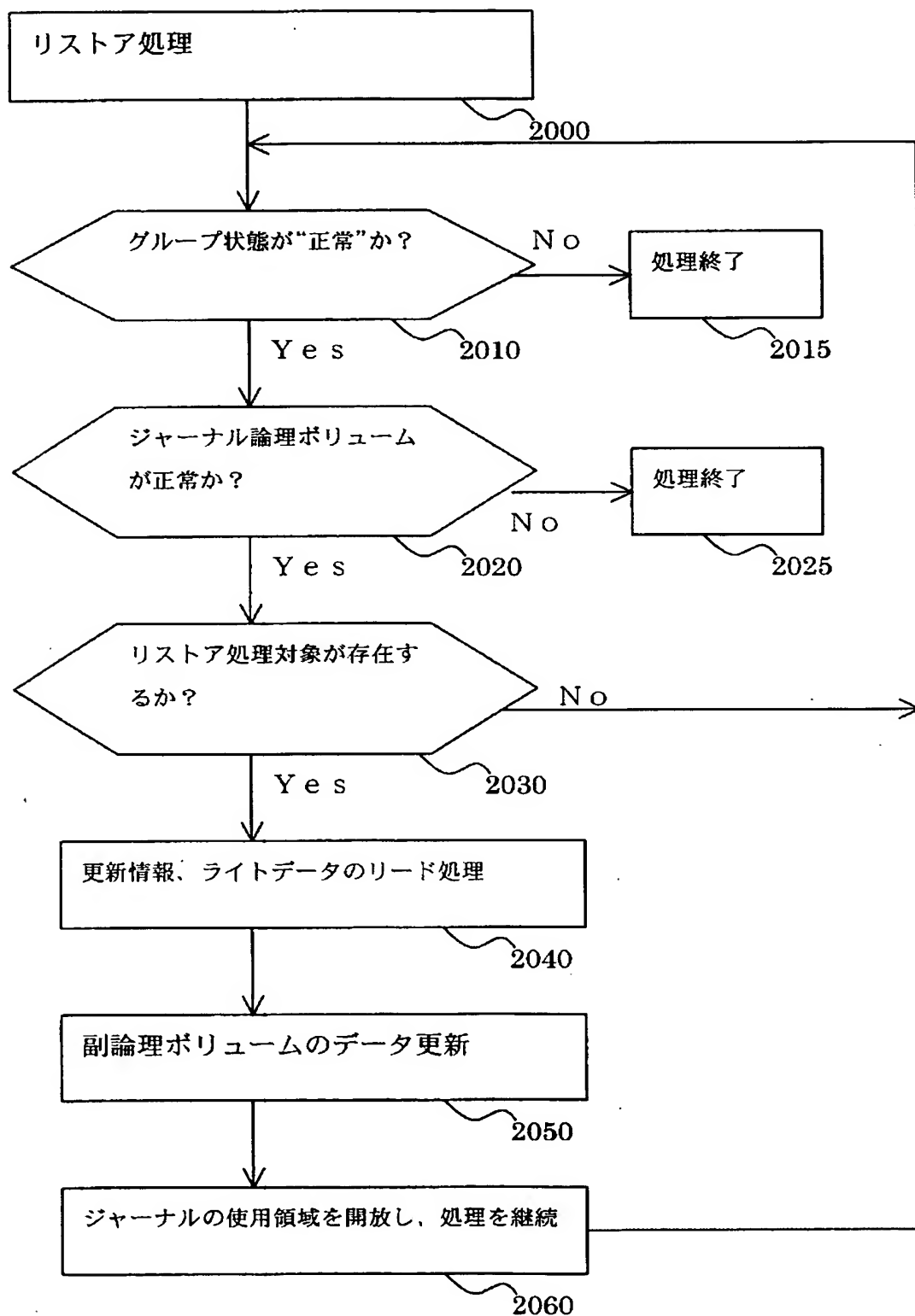
【図 19】

図 19



【図 20】

図 20



【図 21】

図 21

設定項目	設定値例
ライト命令を受信した時刻	1999/3/17 22:20:10
グループ番号	1
更新番号	4
ライト命令の論理アドレス	論理ボリューム番号: 1 論理ボリュームの記憶領域の先頭からの位置: 700
ライトデータのデータサイズ	300
ライトデータを格納したジャーナル論理ボリュームの論理アドレス	論理ボリューム番号: 4 論理ボリュームの記憶領域の先頭からの位置: 1500

310 更新情報

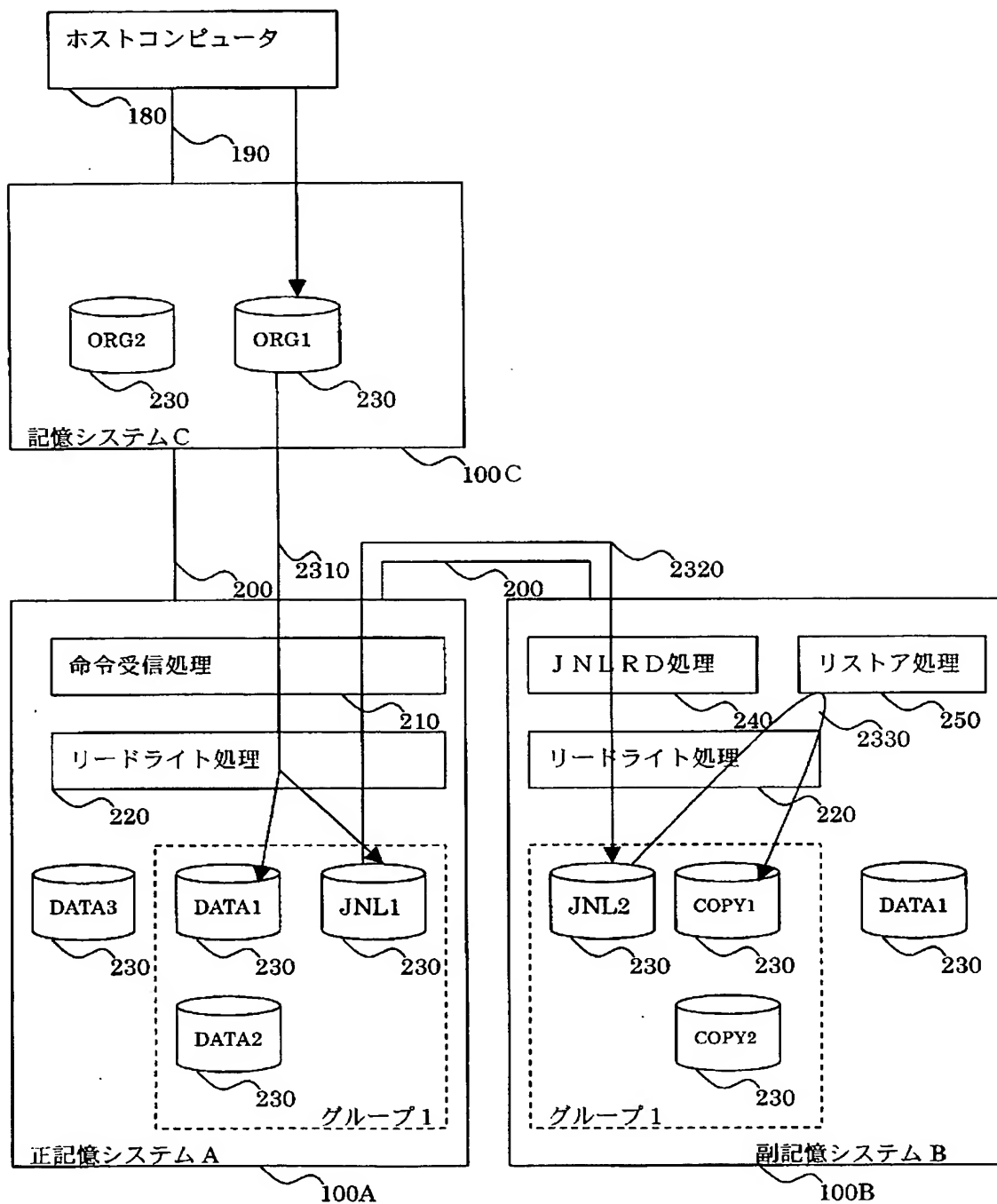
【図 22】

図 22

設定項目	設定値例
ライト命令を受信した時刻	1999/3/17 22:22:10
グループ番号	1
更新番号	4
ライト命令の論理アドレス	論理ボリューム番号: 1 論理ボリュームの記憶領域の先頭からの位置: 800
ライトデータのデータサイズ	100
ライトデータを格納したジャーナル論理ボリュームの論理アドレス	論理ボリューム番号: 4 論理ボリュームの記憶領域の先頭からの位置: 2200

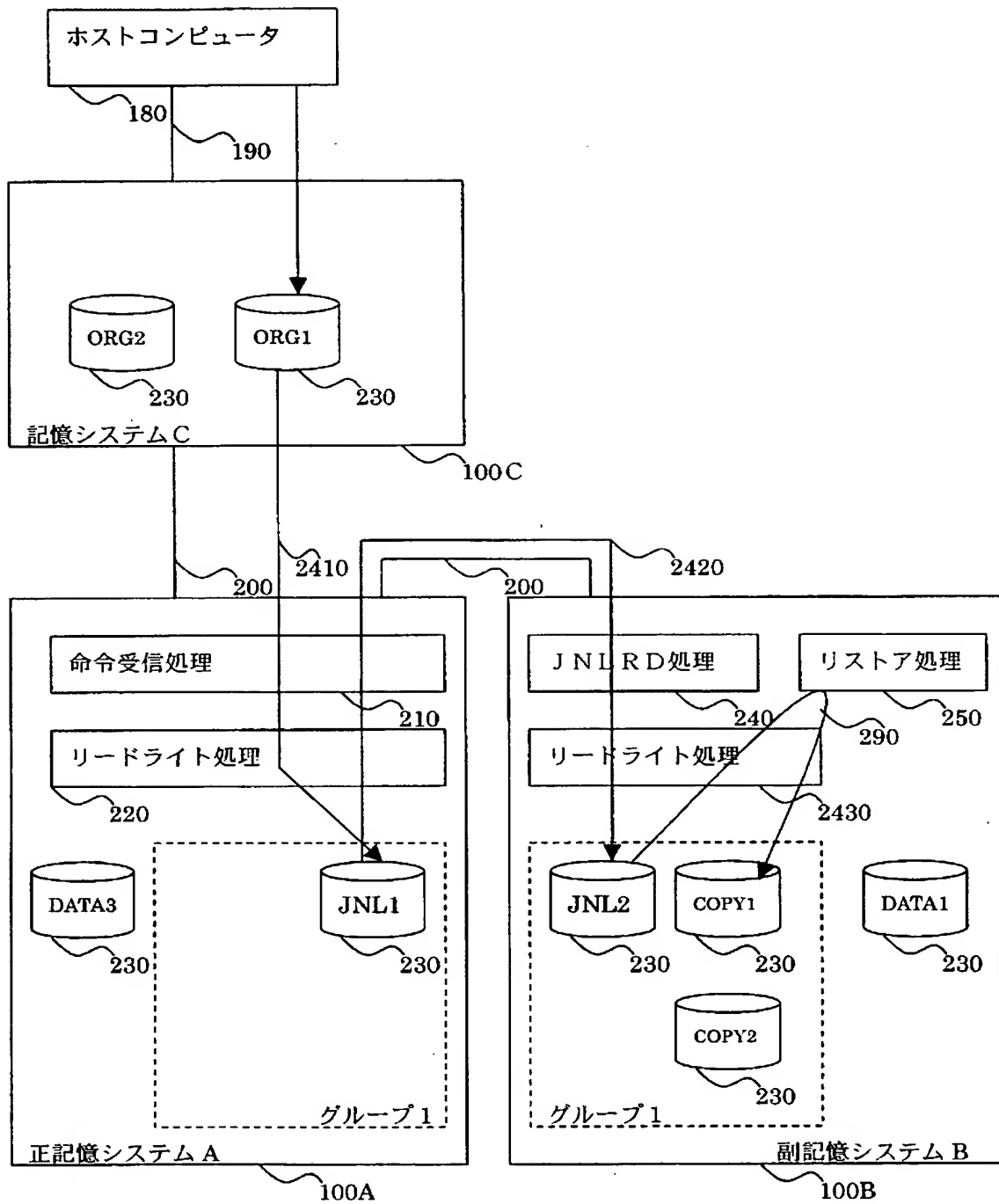
【図 23】

図 23



【図 24】

図 24



【書類名】 要約書

【要約】

【課題】

【解決手段】

第一の記憶システムは、第一の記憶システムに格納されたデータの更新に関する情報をジャーナルとして格納する。ジャーナルは、具体的には、更新に用いられたデータのコピーと更新時のライト命令等の更新情報とによって構成される。さらに、第二の記憶システムは、第一の記憶システムと第二の記憶システム間の通信線を介して、前記ジャーナルを取得する。第二の記憶システムは、第一の記憶システムが保持するデータの複製を保持しており、前記ジャーナルを用いて、第一の記憶システムでのデータ更新順に、第一の記憶システムのデータと対応するデータを更新する。

【選択図】 図 1

認定・付加情報

特許出願の番号	特願 2003-183734
受付番号	50301071602
書類名	特許願
担当官	第七担当上席 0096
作成日	平成15年 6月30日

<認定情報・付加情報>

【提出日】 平成15年 6月27日

特願 2 0 0 3 - 1 8 3 7 3 4

出 願 人 履 歴 情 報

識別番号 [0 0 0 0 0 5 1 0 8]

1. 変更年月日	1 9 9 0 年 8 月 3 1 日
[変更理由]	新規登録
住 所	東京都千代田区神田駿河台 4 丁目 6 番地
氏 名	株式会社日立製作所